Genetic and developmental divergence in the neural crest programme between cichlid fish species

Aleksandra Marconi*[,1,†], Grégoire Vernaz*[,2], Achira Karunaratna[1], Maxon J. Ngochera[3], Richard Durbin[4] & M. Emília Santos[1,†]

[1]Department of Zoology, University of Cambridge, United Kingdom;
[2]Zoological Institute, University of Basel, Switzerland;
[3]Senga Bay Fisheries Research Center, Malawi Fisheries Department, P.O. Box 316, Salima, Malawi;
[4]Department of Genetics, University of Cambridge, United Kingdom.

*These two authors contributed equally to this work.

[†]Correspondence to: am2510@cam.ac.uk & es754@cam.ac.uk

ORCID ID

AM: 0000-0001-8481-9336; **GV**: 0000-0001-8942-2370;
EMS: 0000-0003-3158-7935; **RD:** 0000-0002-9130-1006
AK: 0009-0006-2088-6816

Abstract

Neural crest (NC) is a vertebrate-specific embryonic progenitor cell population at the basis of important vertebrate features such as the craniofacial skeleton and pigmentation patterns. Despite the wide-ranging variation of NC-derived traits across vertebrates, the contribution of NC to species diversification remains underexplored. Here, leveraging the adaptive diversity of African Great Lakes' cichlid species, we combined comparative transcriptomics and population genomics to investigate the evolution of the NC genetic programme in the context of their morphological divergence. Our analysis revealed substantial differences in transcriptional landscapes across somitogenesis, an embryonic period coinciding with NC development and migration. This included dozens of genes with described functions in the vertebrate NC gene regulatory network, several of which showed signatures of positive selection. Among candidates showing between-species expression divergence, we focused on teleost-specific paralogs of the NC-specifier *sox10* (*sox10a* and *sox10b*) as prime candidates to influence NC development. These genes, expressed in NC

cells, displayed remarkable spatio-temporal variation in cichlids, suggesting their contribution to inter-specific morphological differences. Finally, through CRISPR/Cas9 mutagenesis, we demonstrated the functional divergence between cichlid *sox10* paralogs, with the acquisition of a novel skeletogenic function by *sox10a*. When compared to the teleost models zebrafish and medaka, our findings reveal that *sox10* duplication, although retained in most teleost lineages, had variable functional fates across their phylogeny. Altogether, our study suggests that NC-related processes – particularly those controlled by *sox10*s – might be involved in generating morphological diversification between species and lays the groundwork for further investigations into mechanisms underpinning vertebrate NC diversification.

Introduction

The remarkable diversity and complexity of craniofacial structures, pigmentation patterns, and social behaviours within vertebrates is a testament to their outstanding capacity to adapt and exploit a wide range of ecological niches. Much of this phenotypic diversity is intimately connected with the emergence of the neural crest (NC) (Donoghue et al., 2008; Gans and Northcutt, 1983). This embryonic multipotent cell population arises from the dorsal portions of the neural tube and then migrates extensively to finally differentiate into a remarkable range of cell types and tissues, including neurons and glia, pigment cells, craniofacial cartilage and bone, among others (Brandon et al., 2023; Bronner and Simões-Costa, 2016). These diverse cell lineages later assemble to form complex pigmentation patterns in fish, amphibians and birds, as well as divergent head structures, such as fish jaws, bird beaks or mammalian horns (Eames and Schneider, 2005; Elkin et al., 2023; Jheon and Schneider, 2009; Nasoori, 2020).

NC has been primarily studied in the context of its origin, development and function, including developmental disorders involving its derivatives (neurocristopathies) (Bolande, 1997). Studies in model organisms have revealed that the gene regulatory networks (GRNs) and developmental processes governing NC specification, migration and differentiation are highly conserved across distantly related species (Bronner and Simões-Costa, 2016; Simões-Costa and Bronner, 2015). This remarkable macroevolutionary conservation of the NC programme raises key questions about its evolvability and its potential contribution to the origins of vertebrate diversity. Surprisingly, the role that NC cells may play in the the evolution of species-specific traits (i.e. at the population or species level) remains largely unexplored (Abzhanov et al., 2004; Brandon et al., 2023; Donoghue et al., 2008; Kratochwil et al., 2015; Powder et al., 2014). This is despite the rapid and extensive diversification of NC-derived structures - a hallmark of adaptive radiations of multiple vertebrate clades. Striking examples include the diversification of cranial shapes of *Anolis* lizards, beak morphologies in Darwin's finches and, perhaps most spectacularly, the craniofacial skeletons and

colour patterns of cichlid fish radiations in the Great African Rift Lakes (Abzhanov et al., 2006; Albertson and Kocher, 2006; Sanger et al., 2012; Santos et al., 2023).

Here, to investigate the molecular evolution of NC-related phenotypic diversity in the spectacular Lake Malawi cichlid fishes radiation, we first examined the extent of variation in NC genetic and developmental programmes between two closely related, yet eco-morphologically divergent cichlid species, namely *Astatotilapia calliptera* 'Mbaka' and *Rhamphochromis* sp. 'chilingali'. Both species belong to the Lake Malawi cichlid radiation and are characterised by distinct craniofacial morphologies, body pigmentation, ecologies(Bronner and Simões-Costa, 2016) and diets (Edgley and Genner, 2019; Salzburger, 2018; Santos et al., 2023; Turner, 2007) **(Fig. 1a-b** and **Supplementary Fig. 1)**. Our previous work identified variation in pigmentation and craniofacial shapes at the earliest stages of their overt appearance at post-hatching stages (Marconi et al., 2023). Considering the direct mode of development in cichlids (i.e. without larval stage and metamorphosis, unlike zebrafish (Marconi et al., 2023; Woltering et al., 2018), the variation in these NC-derived traits likely stems from differences in early embryogenesis (Marconi et al., 2023) (**Supplementary Fig. 1a-d**).

To test this hypothesis, we focused on the interspecific comparison at earlier embryonic stages concomitant with NC cell specification, migration and onset of their differentiation (Rocha et al., 2020a) (**Fig. 1c-d**). Using whole-transcriptome time-series sequencing data, we uncovered substantial variation in coding and non-coding gene expression throughout NC development between the two species (**Fig. 1d-e** and **Supplementary Fig. 2e-g**) and included divergence in expression levels and temporal trajectories of dozens of genes with key functions within the teleost NC-GRN (Rocha et al., 2020a). Moreover, we show that several of these crucial genes are also associated with signatures of divergent positive selection between species, potentially contributing to species-specific phenotypes. We then focused on two SRY-box transcription factor 10 (*sox10*) paralogs of a key NC specifier, namely *sox10a* and *sox10b*, and showed that they both arose during teleost-specific whole genome duplication and exhibit prevalent inter-specific expression variation

throughout NC development. These results suggest potential contribution of *sox10* paralogs, and NC development more broadly, to species differences. Finally, we provide experimental evidence that *sox10a* function is essential for craniofacial skeletal development, indicating a novel role in cichlids that has not been described in any teleost to date. Taken together, our study reveals that *sox10* paralogs followed divergent functional evolution across the teleost phylogeny, including gene loss (zebrafish), subfunctionalisation (medaka) and neofunctionalization (cichlids). We propose that the expansion of genetic toolkit associated with neural crest development during genome duplication, subsequent lineage-specific divergence of paralogous genes, including the acquisition of novel functions, and regulatory and transcriptomic evolution in cichlids, may have collectively contributed to the extensive morphological diversification in this clade. Our results highlight cichlids as a unique teleost system to investigate the developmental and genetic underpinnings of adaptive phenotypic evolution.

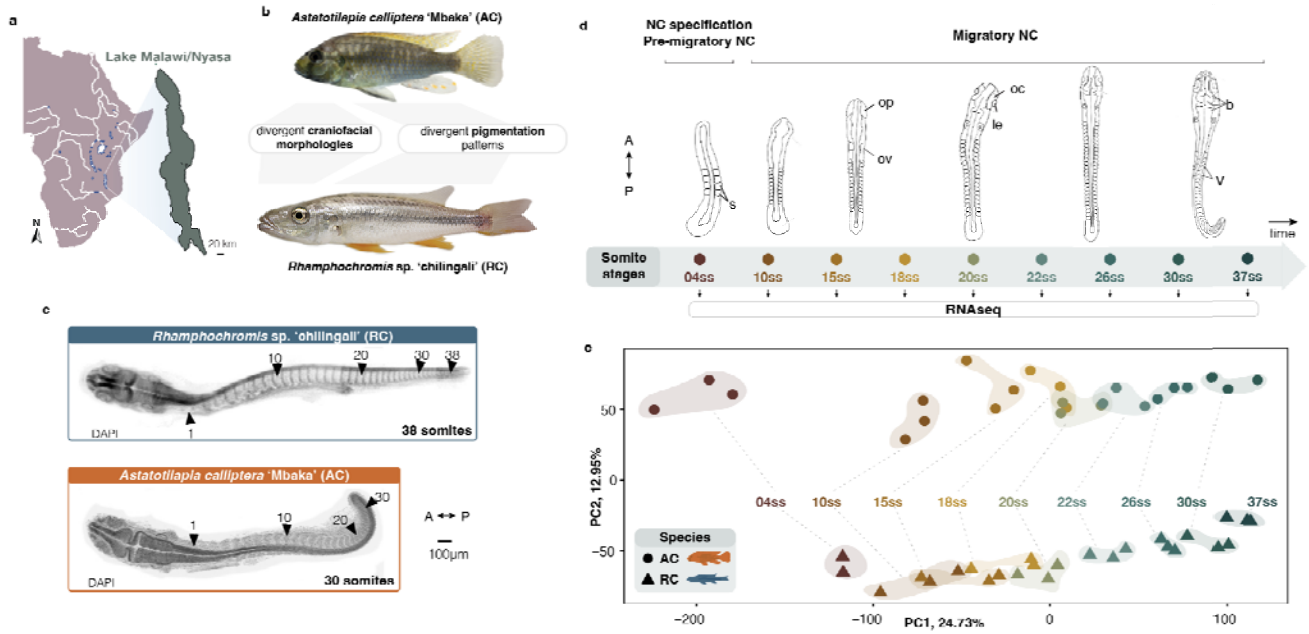


Figure 1. Divergence in whole-embryo transcriptomic trajectories during early embryogenesis and neural crest development between two morphologically distinct Lake Malawi cichlids. a, Geographical map of Lake Malawi/Nyasa. **b**, Cichlid species part of this study exhibit distinct neural crest (NC)-derived craniofacial morphologies and pigmentation patterns (to scale). AC has intermediate phenotype of a generalist feeder with variable melanic patches comprising features of both bars and stripes on a yellow-grey background, whereas RC has a flattened head and elongated jaws typical of a piscivore with silvery colouration and dark horizontal stripes (See Supplementary Fig. 1). **c,** AC and RC exhibit different total somite numbers upon completion of somitogenesis (30 in AC, 38 in RC). **d,** The stages of cichlid somitogenesis (expressed as somite stages, ss) examined in this study and collected for RNA sequencing range from the early stages of NC specification (4ss) through migratory NC (10-12ss onwards) and its differentiation during late somitogenesis stages, concluding at 30 and 38ss in AC and RC, respectively (n = 3 biological replicates per ss). **e,** Principal Component Analysis (PCA) of whole transcriptome samples reveals significant ontogenic (PC1) and species-specific (PC2) clustering. Each data point corresponds to a single replicate embryo. A - anterior, AC - *Astatotilapia calliptera* 'Mbaka', dpf - days post-fertilisation, le - lens, oc - optic cup, op - optic primordium, ov - otic vesicle, P - posterior, RC - *Rhamphochromis* sp. 'chilingali', s - somites, ss - somite stage, st - stage, b - tri-partite brain, V - V-shaped somites. Map (**a**) modified from d-maps.com.

Results

Genes involved in NC development show species-specific shifts in transcriptomic trajectories

Our previous work has shown that phenotypic divergence between cichlid species in NC-derived craniofacial skeleton and body pigmentation is first observed at the appearance of differentiated cartilage and pigment-bearing cells at early post-hatching stages, respectively (Marconi et al., 2023) (**Supplementary Fig. 1c-d**). Given that cichlids do not undergo metamorphosis and develop directly from embryo to adult (Marconi et al., 2023; Woltering et al., 2018), we hypothesise that variation in processes of NC development occurring early in ontogeny might constitute an important contributor to the adult morphological divergence between these species.

To first investigate the potential role of gene expression variation in driving divergent NC-derived phenotypes, we performed comparative transcriptome profiling (RNAseq) of whole embryos of *Astatotilapia* and *Rhamphochromis* across somitogenesis. This period of embryonic development coincides temporally with NC development (Rocha et al., 2020a) (**Fig. 1c**). In total, 32.49 ± 2.5 Mio paired-end 150bp-long reads were generated for each sample (three biological replicates per somite stage, ss), and then aligned against the *Astatotilapia calliptera* reference genome to quantify gene expression **(Supplementary Table 1** and see **Materials and Methods)**. Principal component analysis (PCA) of protein-coding transcriptomes revealed that gene expression in these species is primarily dictated by ontogeny (i.e. their developmental age in ss; PC1, 24.73%), followed by species (PC2, 12.95%; **Fig. 1e** and **Supplementary Fig. 1e**). Conversely, the expression of non-coding transcripts and transcribed transposable elements (TEs) is primarily clustered by species and then by ontogeny (**Supplementary Fig. 1f-h**), in line with previous reports of higher evolutionary rates associated with non-coding genes in cichlids (El Taher et al., 2021).

We then performed differential gene expression analysis across somitogenesis stages to identify gene candidates showing species- and time-specific transcriptional patterns. In total, 12,611 differentially expressed genes (DEGs; $p<0.05$) with >1.5-fold expression difference in at least one pairwise comparison were identified (**Fig. 2a, Supplementary Tables 2-3** and **Materials and Methods**). Of these, 14.7% (n=1,857) lacked assigned names in the current assembly *A. calliptera* genome (Ensembl 108), likely representing genes without zebrafish orthologs or novel genes and were excluded from downstream analyses. The remaining DEGs were then classified into seven distinct clusters based on unbiased grouping according to their expression patterns **(Fig. 2a** and **Supplementary Fig. 1i)**. While four of these clusters displayed consistent high or low gene expression in one species across all somite stages (clusters 2, 3, 5 and 6; accounting for 51.4% of all DEGs), the other clusters (1, 4 and 7) showed species-specific temporal shifts in gene expression, possibly linked to the temporal differences (heterochronies) during somitogenesis between these species (Marconi et al., 2023). Each gene expression cluster was significantly enriched for specific Gene Ontology (GO) categories associated with functions ranging from transcription regulation to metabolic and developmental processes (**Fig. 2b**). Notably, clusters 4 and 7, displaying temporal shifts in gene expression dynamics between the two species during early and late somitogenesis, were significantly enriched for genes with functions related to neural crest cell differentiation (**Fig. 2 a-b** and **Supplementary Fig. 1i**). Altogether, the considerable divergence in expression dynamics between species over developmental time (e.g. temporal shifts) may imply variation in multiple developmental processes during embryogenesis, including those involving NC cells. Combined with the diverse repertoire of NC derivatives, these results implicate the potential role of this cell population in the divergence of species-specific traits.

Signatures of positive selection are associated with NC-related genes

We then sought to assess if DEGs, and in particular NC-related genes, were potentially diverging between species by conducting genome-wide scans for regions under positive selection. Through extended haplotype homozygosity (xp-EHH, Gautier and Vitalis, 2012) scans between wild

Astatotilapia and *Rhamphochromis* populations (43-45 whole genomes per species), we identified 154 regions showing significant signatures of positive selection (**Fig. 2c; Supplementary Fig. 2a**, **Supplementary Table 4, Materials** and **Methods**). Altogether, 74 DEGs were located near or within these putative islands of selection **(Fig. 2d**, **Supplementary Fig. 2b-c;** see **Materials and Methods**) with functions ranging from cell differentiation, signal transduction to metabolic pathways, and included several unannotated novel genes (**Supplementary Fig. 2e**).

Among the annotated DEGs located in the most extreme outlier regions were *fgf8a* (cluster 7), *kcnk18* (cluster 1) and *tspan37* (cluster 5), which exhibited significant expression differences between the two species during the early and mid-phases of somitogenesis (**Supplementary Fig. 2b-c**). *tspan37* is an integral membrane protein involved in cellular signalling, while *kcnk18* encodes a potassium channel expressed in the brain and eye of hatchling and larval zebrafish (Rahm et al., 2014). To test whether these genes were involved in NC development and characterise their expression patterns, we performed *in situ* Hybridisation Chain Reaction (HCR) (Choi et al., 2018). Expression of *kcnk18* and *tspan37* was not detected in whole mount embryos at the stages of differential expression, perhaps due to overall low expression levels **(Supplementary Fig. 2c).** Furthermore, *fgf8a*, albeit known for its diverse roles during embryogenesis, including in chondrogenesis of the NC-derived cranial skeleton (Gebuijs et al., 2019), showed expression only in the developing brain and posterior-most notochord during somitogenesis in the examined cichlids **(Supplementary Fig. 3)**, consistent with findings in zebrafish. This limited expression pattern suggests an unlikely role in craniofacial development and patterning at this stage. Altogether, the top divergent outlier genes are not likely to play a role in NC-derived divergence between these two cichlid species. However, NC-related DEGs (identified based on their GO annotation, **Supplementary Table 2, Materials and Methods**) belonging to cluster 6 and showing overall lower expression in *Rhamphochromis* throughout somitogenesis **(Supplementary Fig. 1i)** displayed significant enrichment for sites under potential positive selection (**Fig. 2e** and **Supplementary Fig. 2d**). These included the transcription factor *pax3a*, already implicated in cichlid interspecific pigment pattern variation (Albertson et al., 2014), and the cellular nucleic acid-binding *cnbpa,* involved in

craniofacial development in fish (Weiner et al., 2007), among others. Collectively, these findings highlight a strong association between sequence and transcriptional differences between the two species, with several DEGs with functions related to NC processes showing enrichment for sites under potential positive selection.

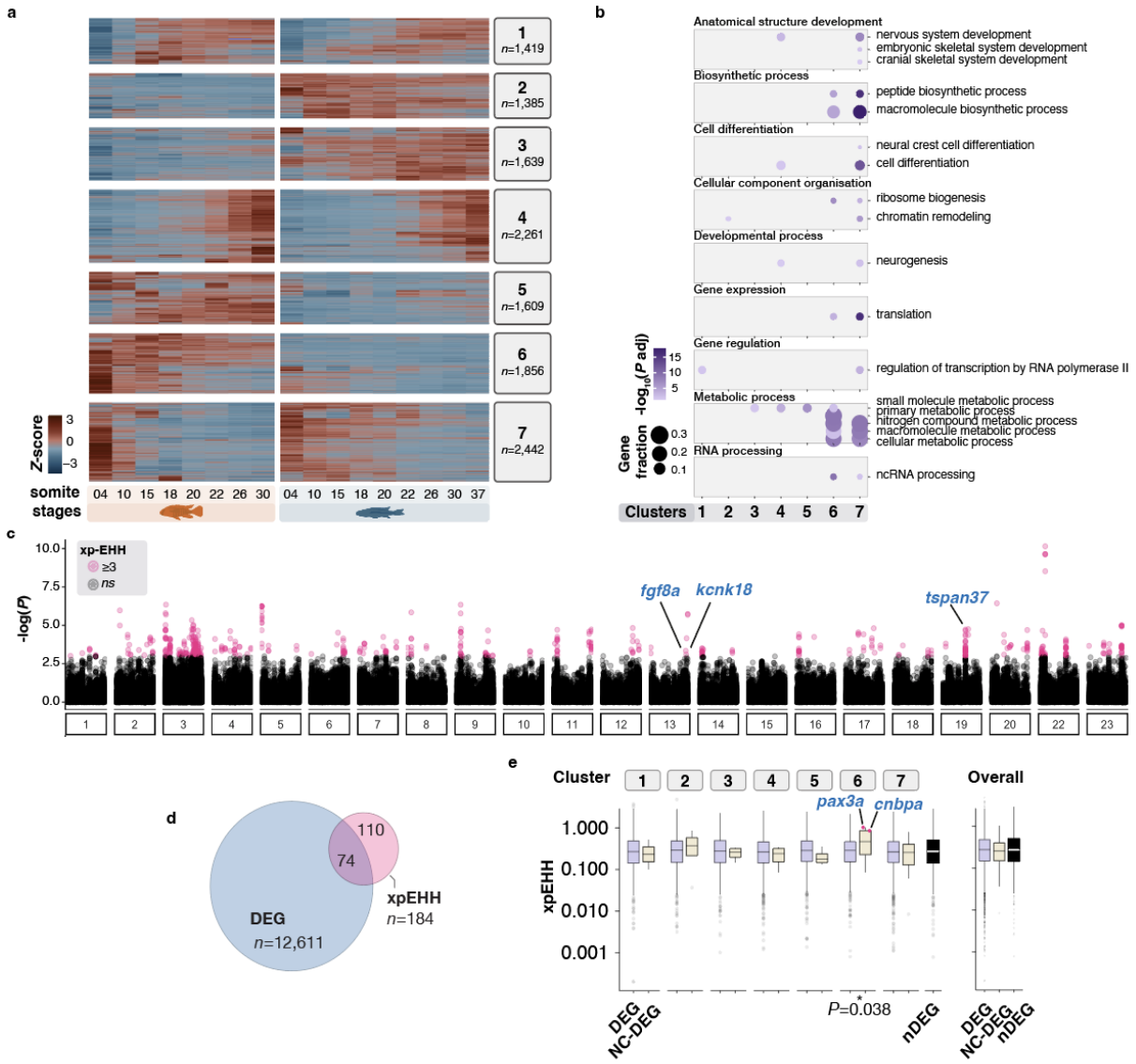


Figure 2. Comparative characterisation of cichlid transcriptomes during somitogenesis and neural crest development. a, Heatmap of all differentially expressed genes (DEGs, n=12,611) between any pairwise somite stage comparisons between species identifies seven clusters of gene expression patterns**. b**, Distinct Gene Ontology categories are significantly enriched in each of the seven clusters of gene expression identified. **c**, Genome-wide scans between *Astatotilapia* and *Rhamphochromis* populations reveal 551 significant SNPs (in pink) showing elevated local

haplotype homozygosity (measured using between-population extended haplotype homozygosity, xpEHH; see also Supplementary Fig. 2b). **d,** Significant SNPs in close proximity (see Materials and Methods) were grouped into 154 islands, which were associated with 74 DEGs. **e**, NC-DEGs within cluster 6 show significantly elevated xpEHH values, in line with signatures of positive selection. DEG, differentially expressed genes; nDEG, genes lacking significant expression differences; NC-DEG, DE genes linked to NC development.

Variation across the NC gene regulatory network is particularly associated with NCC migration

Since our comparative transcriptional and selection analyses highlighted divergence in the genetic programme of NC development between species, we next sought to identify genes, processes and stages associated with early NC ontogeny potentially implicated in the evolution of morphological diversity. To this end, we examined only DEGs with known functions (based on their GO annotation, **Supplementary Table 2**) in NC development and differentiation and its two extensively diversified derivatives, namely pigmentation and craniofacial skeleton **(Fig. 3a)**

The identified NC-DEGs belonged to all tiers of the NC-GRN (Betancur et al., 2010; Sauka-Spengler and Bronner-Fraser, 2008; Simões-Costa and Bronner, 2015) **(Fig. 3a)**. These included genes involved in NC induction, NC specification, NC migration and NC differentiation. This last category included numerous genes associated with the development of different pigment cell lineages, such as melanophores, xanthophores and iridophores (Howard et al., 2021). Multiple genes contributing to the development of the embryonic cranial skeleton, also showed divergence in expression between species as well as four signalling pathways - Bmp, Fgf, Hedgehog and Wnt. Almost 40% of the identified NC-DEGs **(Fig. 3b)** were associated with NCC migration (GO:0001755), and most were differentially expressed at 4ss and 15ss **(Fig. 3b).** These genes (highlighted in red in **Fig. 3c)** exhibited variation in expression levels over time, often displaying large differences in relative expression between species at individual stages, such as 4, 15 and 18ss **(Fig. 3c)**. These findings might reflect divergence in gene expression but also indicate variation in the sizes of cell

populations expressing each of these genes (given that the more cells, the more mRNA will be detected in a bulk approach.

The considerable number of DEGs involved in NC migration and differentiation might result from broad knock-on effects of divergence at early stages (i.e. during NC specification) within the NC programme. We identified three candidate genes – *sox10, prdm1a* and *dicer1* – known to perform multiple functions in NC development, including in specification and migration of NC cells and in differentiation of pigment cells and craniofacial cartilages. *sox10* is a key regulator of NC specification, maintenance, migration and differentiation into multiple cell lineages, primarily neuronal and pigment cells, across vertebrates (Carney et al., 2006; Dutton et al., 2001; Kelsh, 2006). *prdm1a* controls NC cell formation by activating *foxd3* (an early NC specifier gene) and regulating *sox10* in zebrafish (Hernandez-Lagunas et al., 2005; Olesnicky et al., 2010; Powell et al., 2013). *dicer1* is required for craniofacial and pigment cell development, and together with miRNAs, it is involved in the regulation of *sox10* during melanophore differentiation in zebrafish (Weiner et al., 2019).

Both *prdm1a* and *dicer1* were differentially expressed at 4ss, a stage concurrent with NC specification (**Supplementary Fig. 2c**), although *dicer1* was not detected in whole-mount specimens. In *Astatotilapia, prdm1a* was highly expressed in the prechordal plate (anterior-most tip of the neural tube) and at lower levels along the neural tube, whereas in *Rhamphochromis*, it was expressed at lower and uniform levels in the prechordal plate on both sides of the anterior neural tube **(Fig. 3d)**. Considering the positive regulatory role of *prdm1a* on *sox10* (Powell et al., 2013), interspecific expression differences in early NC ontogeny could influence later behaviour of migratory NCCs and their differentiation in lineages regulated by *sox10*, such as pigment cells and cartilage. To test this hypothesis, we next investigated in more detail the expression of *sox10* in cichlid embryos to examine the behaviour of migratory *sox10*-labelled NCCs (Drerup et al., 2009; Dutton et al., 2001).

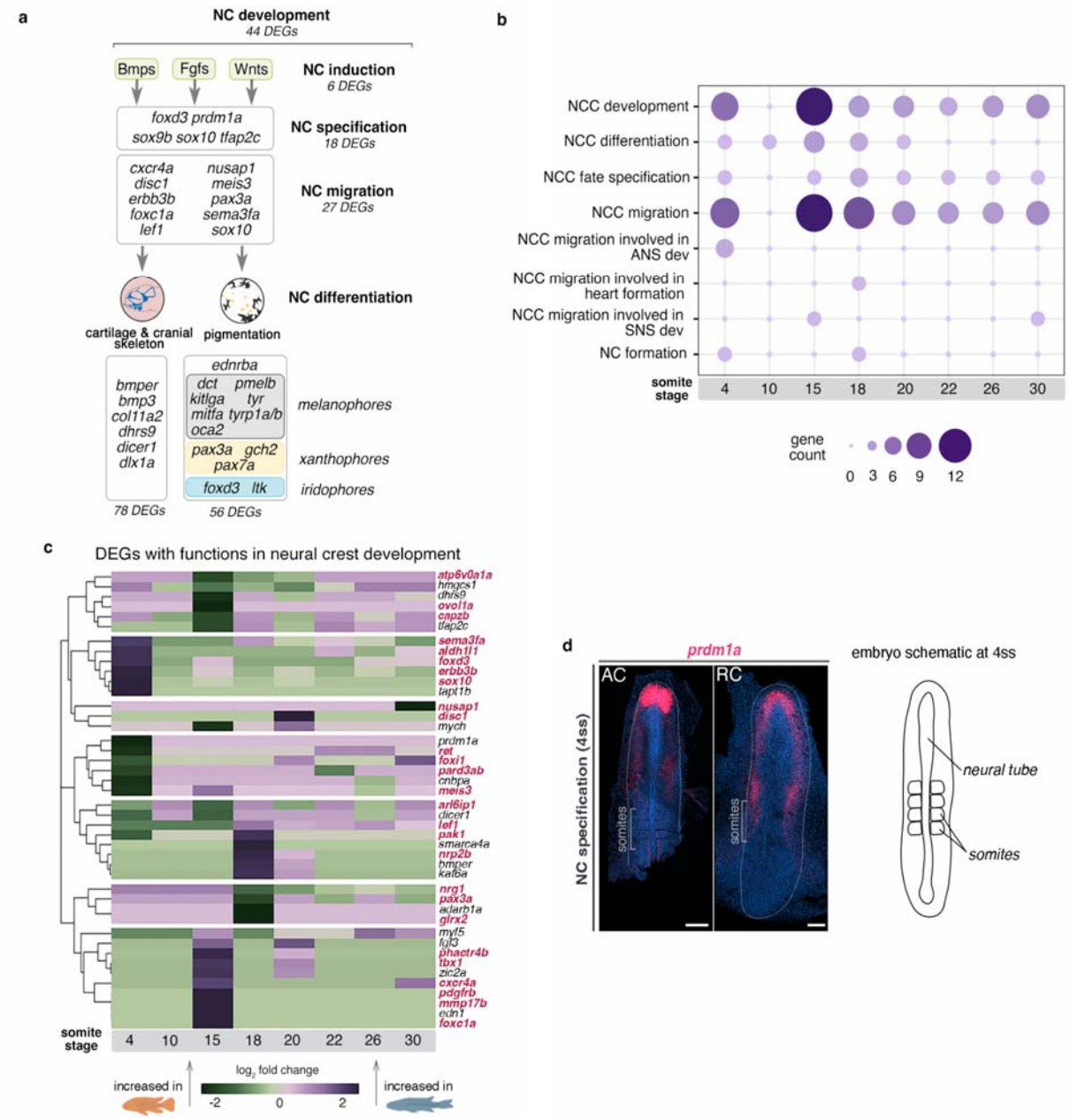


Figure 3. Prevalent transcriptomic variation between cichlids exists across the NC genetic programme. a, identified DEGs belong to different tiers of the teleost NC-GRN (Rocha et al., 2020a), from specification to migration and differentiation, including development of NC-derived pigmentation and craniofacial skeleton. b, distribution of identified NC-DEGs per function along the RNAseq time-course. c, fold changes in expression levels of candidate genes involved in NC development between cichlids. Genes involved in NCC migration highlighted in red. d, in situ HCR image showing prdm1a expression at 4ss, representative of n ≥ 2 per species. Anterior to the top of the figure. AC - *Astatotilapia calliptera* ‘Mbaka’, ANS - autonomic nervous system, NCC - neural crest cells, RC - *Rhamphochromis* sp. ‘chilingali’, SNS - sympathetic nervous system, ss - somite stage. Scale bar = 100 µm.

sox10 paralogs originated in teleost-specific whole genome duplication while _sox10a_ was lost in zebrafish and cavefish

Similar to other members of *soxE* gene family, *sox10* gene is present in two copies (*sox10a* and *sox10b*) in the genomes of cichlids and the majority of other teleosts (Lang et al., 2006; Nagao et al., 2018), with the notable exception of zebrafish – a widely utilised teleost model in biomedical research, which possesses only a single copy, *sox10b* (Braasch et al., 2007; Voldoire et al., 2017). Using recent genomic data of basal teleosts, eels and tarpons (Parey et al., 2023), we confirmed that the *sox10* paralogs indeed originated during the teleost-specific whole-genome duplication event and were subsequently lost in the zebrafish and cavefish lineages **(Fig. 4a)**. Such expansion of the genetic toolkit could provide opportunity for functional divergence, potentially contributing to the process of teleost diversification.

To examine the development of migratory *sox10*-labelled NCCs, we thus focused on both *sox10a* and *sox10b* paralogs. Our transcriptomic profiling revealed that expression of *sox10* duplicates followed similar trajectories over time in both species, with a significant difference in transcript levels between paralogs observed at 10ss (two-way ANOVA and Tukey HSD, $p<0.001$) **(Fig. 4b)**. Further, the *sox10* paralogs were differentially expressed between species across multiple stages of NC development, with both genes showing consistent upregulation in *Rhamphochromis.* Significant fold expression differences in *sox10b* levels were observed at earlier stages compared to *sox10a* and decreased for both genes with developmental time **(Fig. 4c)**. The differences in expression levels of *sox10* paralogs between and within species during embryonic NC development **(Fig. 4)** suggests that the NC developmental programme, and more specifically NC migration, may be divergent between these two closely related species.

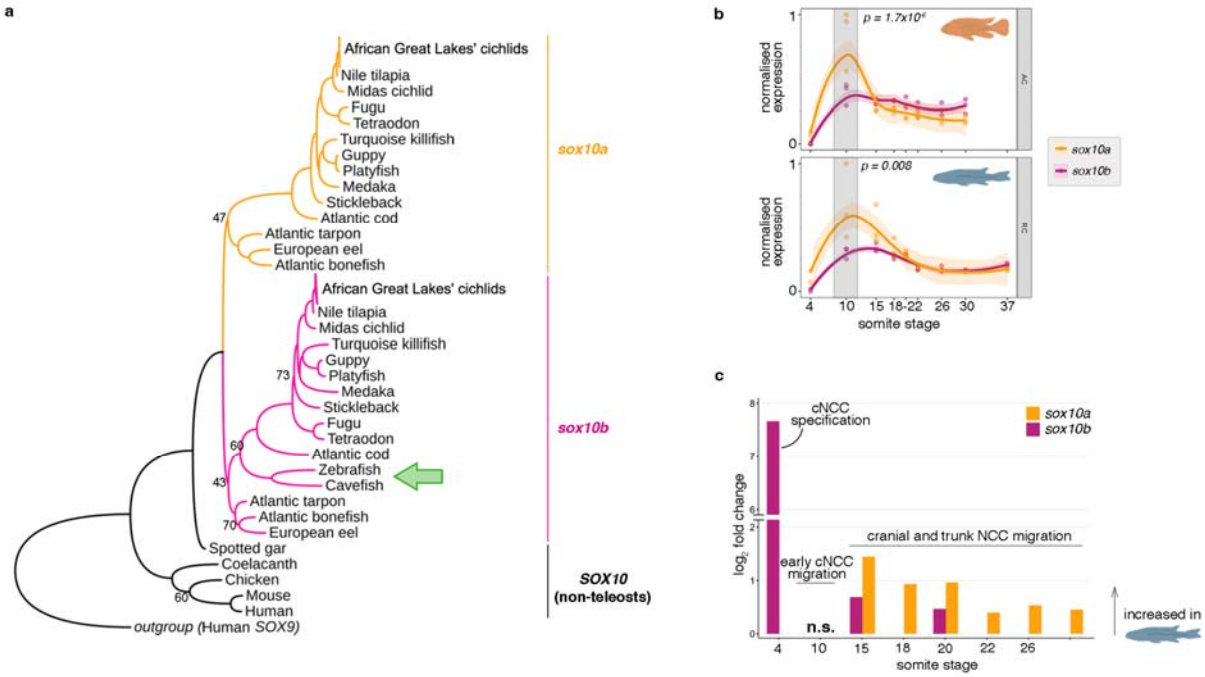


Figure 4. *sox10* paralogs were duplicated in teleosts and are differentially expressed during embryonic development in cichlids. a, The topology of Maximum Likelihood phylogeny of *sox10* paralog coding sequences across vertebrates confirms *sox10* duplication occurred at the root of teleosts, followed by *sox10a* loss in zebrafish and cavefish (green arrow). All significant Bootstrap values, apart from the ones shown (<75). **b,** The expression trajectories of *sox10* paralogs across NC development in *Astatotilapia calliptera* 'Mbaka' (AC) and *Rhamphochromis* sp. 'chillingali' (RC). Shaded bands indicate 95% confidence intervals. Significant differences in normalised gene counts between *sox10* paralogs were observed at 10ss in both examined species (two-way ANOVA and Tukey HSD, $p < 0.01$). **c,** Fold changes in expression levels of *sox10* paralogs between species. Only significant comparisons are shown (P-adj < 0.05). cNCC - cranial neural crest cell, NCC - neural crest cell, n.s. - non-significant.

Differences in expression patterns between *sox10* paralogs suggest divergence in cranial NC development between species

Next, we set out to characterise the precise expression patterns of *sox10a* and *sox10b* in cichlid embryos during the course of NC development. Consistent with findings in medaka (Nagao et al., 2018), the expression of both *sox10* paralogs was generally observed in NC cells at all stages examined, encompassing the processes of NC specification and migration. Furthermore, this

expression was detected across different axial levels, corresponding to distinct NC subpopulations (Rocha et al., 2020b), as well as in the otic vesicle (**Fig. 5**). These findings indicate that both cichlid *sox10* paralogs are likely to perform functions in NC development.

We identified variation in expression between genes within species during cranial NC specification (4ss) which manifested in temporal and spatial aspects common to both examined species. First, *sox10a* was expressed earlier than *sox10b* and concomitant with *foxd3* (4ss in **Fig. 5a, Supplementary Fig. 4a-b**). Second, *sox10a* was detected in cells residing bilaterally in the dorsal region of somites and that did not co-express neither *sox10b* nor *foxd3* **(Fig. 5a-a^{iii})**. Notably, expression of *sox10a* in the somitic region was also observed at later stages in both species and localised into the extraembryonic membranes extending laterally from the embryo proper enveloping the yolk **(Fig. 5c)**. We propose that these *sox10a*+ cells may contribute to the solitary pigmented melanophores that populate yolk during somitogenesis, as they first appear on both sides of the somites in the anterior trunk region at mid-segmentation stages, prior to extensive migration. However, the embryonic origin and function of this novel population remains to be elucidated. Major differences were also observed between species, including multiple *sox10a* domains distributed along the anterior-posterior embryo axis at 4ss in *Astatotilapia* but not *Rhamphochromis* **(Fig. 5a)**. These multiple distinct expression domains of *sox10a*, including in the somitic region, and distinct from those of its paralog *sox10b*, have not been reported for any of the *soxE* family genes in other teleost species to date (Nagao et al., 2018; Takamiya et al., 2020; Tsunogai et al., 2021). Our results thus suggest that *sox10a* could have been co-opted to function in both NC and non-NC cells during early somitogenesis in cichlids, whereas *sox10b* expression during NC specification resembles that of other vertebrates (Aoki et al., 2003; Cheng et al., 2000; Southard-Smith et al., 1998).

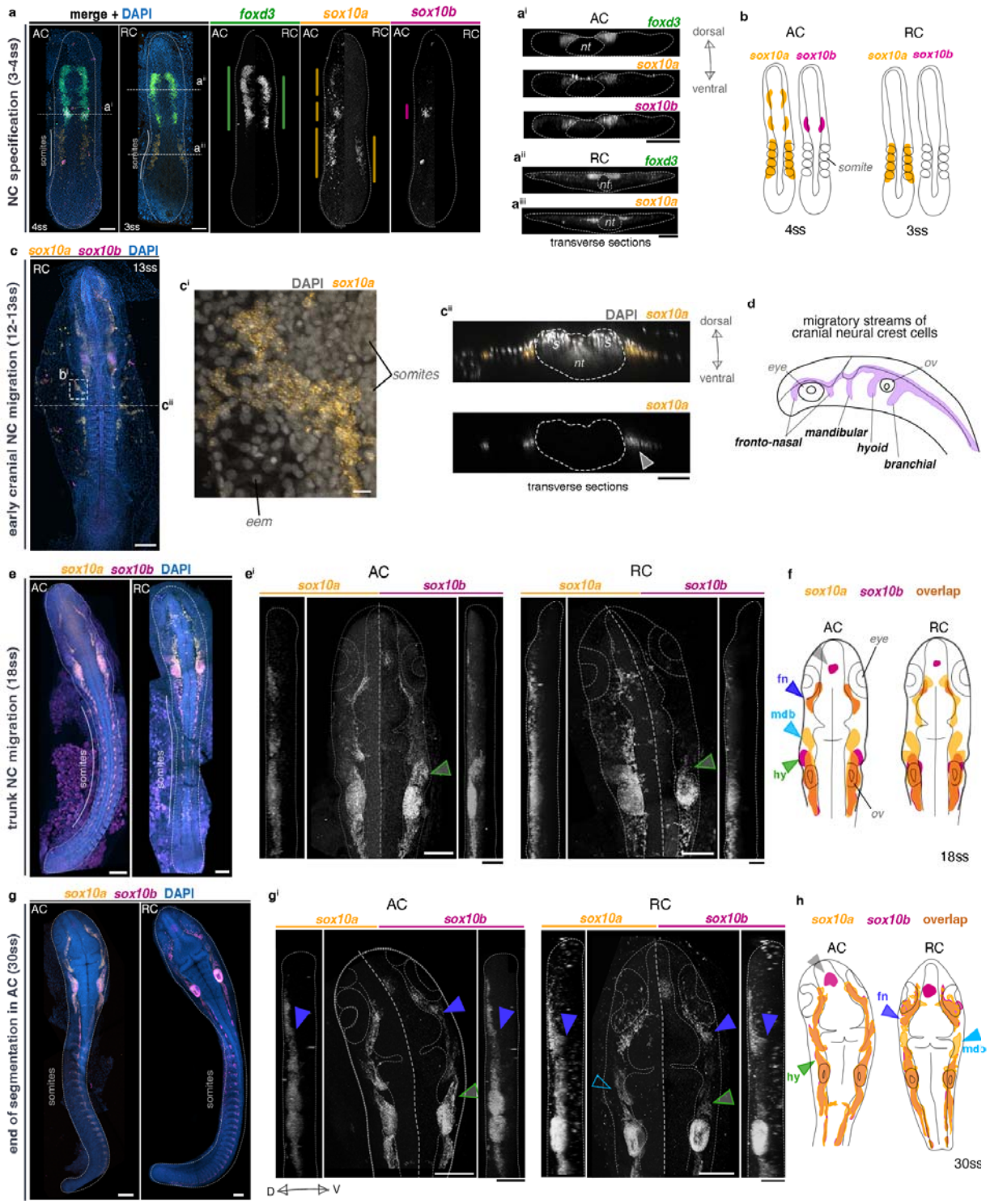
As the development progressed, the variation in expression levels and spatial arrangement of *sox10* paralog domains visibly decreased, leading to largely overlapping patterns during cranial and trunk NC migration in both species (**Fig. 5e-h**)**.** We primarily focused on cranial migration, as differences

in expression patterns between paralogs and between species were most pronounced in that aspect. During cranial migration, NCCs migrate along highly conserved pathways across all vertebrates, forming four main streams: the anterior-most fronto-nasal stream, followed by the mandibular, hyoid and post-otic branchial stream (**Fig. 5d**) (Steventon et al., 2014). Among these, subsets of cells continued to express only one of the paralogs, for example in *Rhamphochromis* at 30ss, the cells migrating ventrally in the mandibular stream **(mdb on Fig. 5g-h)** expressed only *sox10a* (empty blue arrowhead, **Fig. 5h)**, whereas in *Astatotilapia*, cells migrating in the same stream expressed both *sox10a* and *sox10b* **(blue arrowhead, Figure 5h)**. Furthermore, although both paralogs were expressed in the otic vesicles (but not necessarily co-expressed by the same cells, see **Supplementary Fig. 4c**), only *sox10b* was expressed in a group of cells on the dorsal surface of the forebrain, corresponding to oligodendrocytes derived from neural stem cells (Schebesta and Serluca, 2009) **(grey arrowhead, Fig. 5f and h).** Given the conservation of this pattern across vertebrates (Suzuki et al., 2017), these findings suggest that, compared to *sox10a,* cichlid *sox10b* likely performs a broader range of the conserved vertebrate *sox10* roles in fate regulation of some non-NC derived lineages, such as those derived from neural stem cells.

Several differences were also observed between species in the spatial arrangement and migratory behaviour of the cranial NC cell subpopulations labelled by one or both of the paralogs until the end of somitogenesis in *Astatotilapia* (30ss) **(Fig. 5).** The most pronounced divergence involved the extents of their migration into and around head structures (e.g. otic vesicles, eyes**, Fig. 5)**. For instance, at 18ss, *sox10b+* cells migrating in the hyoid stream **(hy on Fig. 5f and h)**, were present further ventral-laterally in *Astatotilapia* compared to *Rhamphochromis* (green arrows on **Fig. 5e**^{i}). This pattern persisted at 30ss **(Fig. 5g**^{i}**)**. In contrast, at the same stage, cells co-expressing *sox10a/sox10b* and migrating in the fronto-nasal stream **(fn on Fig. 5h)** were found further ventrally in *Rhamphochromis* **(purple arrowheads on Fig. 5g**^{ii} **and h)**.

In summary, cichlid *sox10a* and *sox10b* were expressed by overlapping and divergent subsets of cranial neural crest cells migrating along the stereotypical pathways. Notably, we observed fine-

scale variation in migratory patterns of all streams, except for branchial. Such spatial differences could potentially lead to divergence in fine-scale patterning of the structures derived from cranial NCCs i.e. craniofacial cartilages and bones, connective tissues and pigment cells (Kague et al., 2012; Schilling and Kimmel, 1994; Wada et al., 2005). Combined with differences in expression levels from whole-embryo RNAseq **(Fig. 4a-b)**, these differences could be also related to size variation of the populations expressing each paralog, especially considering the differences in embryo sizes between these species (Marconi et al., 2023). Taken together, the variation in expression patterns of *sox10* paralogs in both overlapping and distinct domains indicates divergence in NC development between cichlid species and could reflect potential divergence in developmental functions between two genes.



411

412 **Figure 5. Gene- and species-specific variation in embryonic expression of *sox10* paralogs**
413 **during neural crest development in Malawi cichlids.** *in situ* HCR images of *foxd3* and *sox10*
414 paralogs in representative somite stage-matched embryos of AC and RC. a-a^{iii}, early *sox10* paralog
415 expression partially overlaps with *foxd3*, supporting their expression in *bona fide* NCCs, whereas
416 unique somitic domains suggest expression in non-NC cells. B, schematic representation of sox10

paralog expression patterns at NC specification. C-c^{ii}, during mid-somitogenesis, *sox10a* is also expressed in extra-embryonic tissues in both species, here shown in RC. d, schematic representation of four migratory streams of cranial NC in cichlids (based on expression of *sox10*) in lateral view. e-h, *sox10* paralog expression in the cichlid embryonic head shows differences between genes and between species, implicating differences in the migratory patterns of NCCs. Schematic representations show expression overlays, highlighting overlapping and unique *sox10a/sox10b*-expressing cell populations. All images present maximum intensity projections of dorsal or lateral (vertical panels in e^{i} and g^{i}) views of dissected embryos, with anterior towards the top of the figure. Embryos and sections are outlined in white. Colour-coded vertical bars in a represent the AP extent of expression of each presented gene. Embryos presented in d and e representative of n ≥ 3. AC - *Astatotilapia calliptera* 'Mbaka', D - dorsal, NC - neural crest, nt - neural tube, RC - *Rhamphochromis* sp. 'chilingali', ss - somite stage, V - ventral. Scale bar = 50 µm, 10 µm in c^{i}.

Genome editing reveals functional divergence between *sox10* paralogs, including a novel craniofacial skeletal function of *sox10a* in cichlid fishes

Given the extent of divergence between *sox10a* and *sox10b* expression, we set out to characterise their function in cichlids. *sox10* paralog function remains uncharacterised beyond the zebrafish (*Danio rerio*) and medaka (*Oryzias latipes*) model systems. In zebrafish, *sox10b* function is limited to pigmentation and neural derivatives (Kelsh, 2006), while in medaka both paralogs have redundant functions in pigmentation development (Nagao et al., 2018). To test whether *sox10* genes perform divergent roles in cichlids, we deployed CRISPR/Cas9 system in *Astatotilapia* to induce indel mutations in the coding sequence (exon 1) of *sox10a* and *sox10b* in turn **(Fig. 5, Supplementary Fig. 5-6)**.

From day 6-7 post-injection (st. 18-19), coinciding with the main stages of craniofacial cartilage development and patterning in this species (Marconi et al., 2023), craniofacial malformations were observed in *sox10a* CRISPR mosaic embryos. Neurocranial and craniofacial deformities were prevalent among injected embryos (**Fig. 5a**, n = 21/76 across four clutches, **Supplementary Table 5**) and, while ranging in severity between clutch-mates, these mutants consistently exhibited

flattening of the frontal bones (brain case) (indicated by white dashed lines in **Fig. 5a**), small and bulging, forward-facing eyes as well as protruding, unmoving jaws (white arrowheads in **Fig. 5a**). Alcian Blue stains for cartilage further revealed severely malformed or entirely missing super-orbital cartilages, basihyal, branchial arches and pectoral fins (grey arrowheads and asterisks in **Fig. 5a**). Besides craniofacial abnormalities, *sox10a* mutants at this stage also displayed cardiac and circulatory system defects, reduced black melanophore pigmentation and malformed caudal fin cartilages **(Supplementary Fig. 5)**. The defects in pigmentation were not quantified due to the wide range of severity of cranial deformations, which could have had indirect effects on the pigmentation, for instance due to reduced epidermis surface area for populating chromatophores. The flattened frontal skull slope suggests that the embryonic brains were also likely adversely affected. Mosaic embryos with severe *sox10a*-KO phenotypes **(Fig. 5c** and **Supplementary Fig. 5)** did not survive past 9 dpf (st. 22), suggesting embryonic lethality of a complete KO.

Unlike *sox10*a CRISPR mosaic embryos, *sox10b* mutants did not show any craniofacial cartilage defects at day 7 post-injection (st. 19) nor later, but instead had significantly reduced melanophore pigmentation on the dorsal head region (**Fig. 5b-c,** n = 13/77 across three clutches, **Supplementary Table 5**), the first body area consistently populated by all three differentiated pigment cell types (Marconi et al., 2023). The development of other pigment cell lineages (i.e. reflective iridophores and yellow xanthophores) appeared unaffected by the induced mutations, with no noticeable differences in development of the flank pigmentation at 12 dpf (st. 24) **(Supplementary Fig. 6)**. Despite observed *sox10b* expression in otic vesicles and oligodendrocytes **(Figure 5e-g)** and a known role of zebrafish *sox10b* in glial development (Carney et al., 2006), we did not identify any discernible phenotypes in KO fish involving these tissues.

These functional analyses provide compelling evidence for divergent roles of cichlid *sox10a* and *sox10b* in the development of the neural crest and its derivatives, cartilage and pigment cells (melanophores), respectively. Although we observed a level of potential functional redundancy between paralogs in the differentiation of pigment lineages which requires further investigation, we

also uncovered a novel and pivotal role for *sox10a* in the formation of cranial skeleton (neurocranium and craniofacial cartilages) - a function so far only described in cichlids **(Fig. 7)**. Taken together, the differences in *sox10a* and *sox10b* expression and the *sox10a* role in craniofacial skeletal development show that the cranial NC programme is diverging between cichlid species, but whether the *sox10* paralogs are causally associated with this divergence remains to be tested.

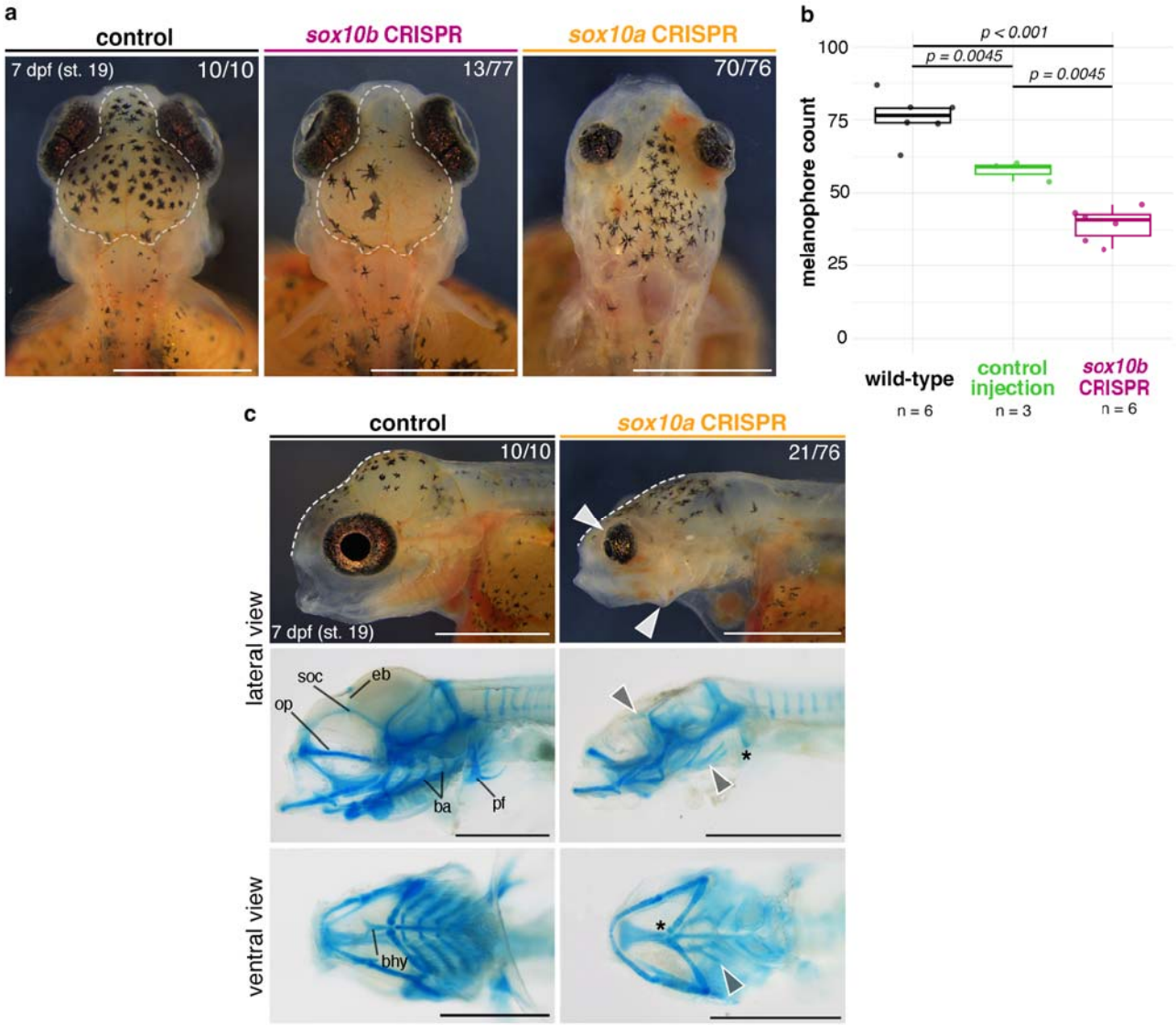


Figure 6. Paralog-specific knockout phenotypes indicate functional divergence of *sox10a* and *sox10b* in cichlids. a, Melanophore pigmentation defects are observed in *sox10b* but not in *sox10a A. calliptera* mutants. **b,** *sox10b*-CRISPR embryos have significantly reduced melanophore

counts at 7 days post-fertilisation compared to WT and control injections (boxplots showing melanophore count within the dorsal head area outlined in a.; *P*-values are shown for Tukey HSD). **c,** Representative craniofacial phenotypes of *sox10a*-CRISPR embryos in live animals and their corresponding cartilage preparations in lateral and ventral views. White dashed lines show the reduced frontal slope of the brain case. Arrowheads and asterisks in panels showing cartilage preparations indicate reduced and missing cartilages, respectively, in the mutants. Fractions indicate the frequency of presented phenotype across all surviving embryos on that day. Data shown are representative of at least three biological replicates per target gene. Ba - branchial arches, bhy - basihyal, dpf – days post-fertilisation, op - orbital process, soc - super-orbital cartilages, pf - pectoral fin, WT - wild-type. Scale bars = 1 mm.

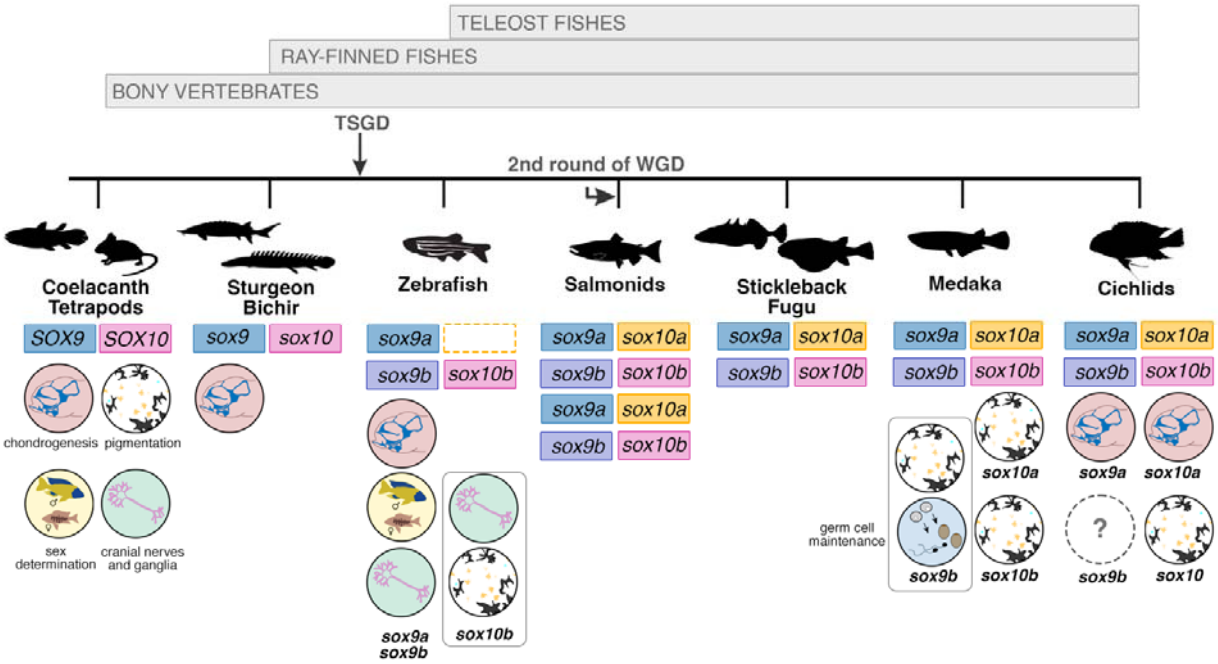


Figure 7. Functional repertoires of *sox10* and *sox9* paralogs across the bony fish phylogeny. Dashed line box indicates gene loss. Bony fish lineage adapted from (Volff, 2005). References for functional analyses are presented in Supplementary Table 6. TSGD - teleost-specific genome duplication; WGD - whole genome duplication. Silhouettes downloaded from http://phylopic.org.

Discussion

The genetic programme orchestrating development of the neural crest is remarkably conserved across vertebrates, despite NC-derived structures constituting some of the most diverse phenotypic traits, especially among lineages that have undergone adaptive radiation. Our study suggests that neofunctionalization following gene duplication, together with extensive transcriptomic divergence during early NC development, may have collectively contributed to the morphological diversification of NC-derived traits, including pigmentation and craniofacial shapes, in East African cichlids. On a larger evolutionary scale, we report a rare example of taxon-specific, divergent evolutionary trajectories of paralogous genes originating from a single genome duplication event in vertebrates, with *sox10* paralogs showing different fates in different teleost taxa.

Inter-specific divergence in transcriptional landscape and signature of positive selection in NC-related genes

Recent studies have highlighted that transcriptional evolution among closely-related species and across tissues often underlies phenotypic diversity (Cardoso-Moreira et al., 2019; El Taher et al., 2021). Our findings support and expand upon this concept by uncovering distinct transcriptomic dynamics during somitogenesis between two closely related yet eco-morphologically distinct Malawi cichlid species, with notable differences between coding genes and non-coding and transposable element transcripts. Firstly, we observed that long non-coding genes and TEs displayed predominantly species-specific expression trajectories and presumably higher evolutionary rates compared to protein coding genes. These findings are particularly intriguing and warrant further work, as non-coding genes and TEs have the potential to contribute to phenotypic evolution through multiple mechanisms, including generation of genetic variation, modification of gene regulation and the alteration of genomic architecture (Almeida et al., 2022; Wells and Feschotte, 2020). Although recent case studies have highlighted the roles of TEs in cichlid diversification (Brawand et al., 2014), understanding of their activity during embryonic development remains very limited.

In addition to the non-coding transcriptome, many protein-coding genes, including those involved in the NC development and its derivatives, also exhibited significant variation in expression trajectories between the two species. Notably, differences in expression of many genes could be explained by simple temporal shifts in timing of gene expression relative to somite stage, which in turn can be attributed to differences in developmental timing (i.e. heterochrony) between species during somitogenesis (Marconi et al., 2023). These results provide evidence that variation in developmental timing, and consequently altered gene expression dynamics, contribute to species divergence in this clade.

Using combination of genomic and transcriptomic data, we identified 74 DEGs with signatures of divergent positive selection between species, with the most extreme outliers related to cellular functions, such as *fgf8a* (secreted signalling molecule)*, tspan37* (regulator of cellular signalling) and *kcnk18* (potassium channel protein), and involved in development of several organs and systems, such as brain, eye and nervous system (Gebuijs et al., 2019; Rahm et al., 2014). The diversity of their functions during development makes it challenging to identify a specific phenotype under selection as multiple traits could be involved simultaneously. Further work could examine the functions of these genes in cichlid embryonic development and verify their expression in adult tissues to provide some insights into these results.

Furthermore, our study identified dozens of DEGs with known NC functions, ranging from specification to differentiation into pigment and cartilage cell lineages. Several of these genes have previously been associated with NC-trait variation in cichlid fishes (Albertson et al., 2014) and appear to evolve under divergent positive selection, implying that they might be playing a role in the adaptive evolution of cichlids. The overrepresentation of DEGs involved in NCC migration, including many cell intrinsic (e.g. transcription factors) and extrinsic factors (e.g. signalling molecules), further underscores the causal role that variation in NC migration might play in the emergence of novel NC

phenotypes and species differences in cichlids, as posited by studies across vertebrates (Fish et al., 2014; Powder et al., 2014; Tucker and Lumsden, 2004).

Differences in NC migration between cichlid species potentially associated with trait divergence

Our comparative analysis in two eco-morphological divergent Malawi cichlids revealed that *sox10* paralogs – master regulators within the NC programme – show species-specific temporal and spatial variation, especially in migratory cranial NCCs. This variation could be in turn linked to pigmentation, craniofacial diversity and potentially differences in other NC-derived cell lineages, such as cranial sensory ganglia, Schwann cells and cardiomyocytes (Sande-Melón et al., 2019; Schilling and Kimmel, 1994).

Specifically, *sox10a* had consistently higher expression in *Rhamphochromis* in somite-stage matched embryos. *sox10a*+ cells in the fronto-nasal stream showed more advanced in ventral migration and earlier expression in the hyoid stream, suggesting that earlier and, potentially, prolonged migration could lead to larger structures, consistent with the prominent jaws of the piscivore compared to moderate phenotype of an omnivore. However, it is important to note that our *sox10a* mutants displayed relatively normal lower jaw cartilages, while the cartilages of branchial arches (originating from NCCs migrating in the branchial stream) and the frontal slope (fronto-nasal stream) were severely affected. These findings align with patterns of *sox10a* expression in both of these streams throughout cranial NC migration.

Further studies involving later stages (e.g. pharyngula until st. 16 when the first cartilages are present, Marconi et al., 2023) and the use of additional molecular markers (e.g. *sox9* genes, *dlx2a,* collagen gene *col2a1a*) are needed to investigate the potential connection between *sox10a*-expressing cells in the embryo, cartilage formation and these affected structures. Future comparative studies on NC development across multiple divergent species, particularly those

characterised by different eco-morphotypes, could provide valuable insights into the connection between NC development, migratory behaviours and adult phenotypes, and how observed differences are associated with their adult adaptive diversity. Nonetheless, our findings highlight the variability of the NC development in closely related cichlids, suggesting unexplored variation in the NC gene regulatory network and highlighting *sox10* paralogs as potential candidates involved in trait diversification.

Evolution of the NC programme following the teleost genome duplication

One consequence of whole genome duplication (WGD) during teleost evolution is the duplication of entire gene networks. Retained genes post-WGD can contribute to functional innovation and evolution through coding sequence changes and rewiring of the regulatory networks controlling gene expression. The latter is particularly relevant for duplicated regulatory genes, which have a higher potential for significant impacts on gene expression and phenotypic effects (Taylor and Raes, 2004). Our phylogenetic analyses confirmed that *sox10* was duplicated during teleost WGD and retained in duplicate in most teleost genomes, except for zebrafish and cavefish. Interestingly, our expression and functional analyses data revealed distinct fates for *sox10* paralogs in cichlids compared to other teleosts. This presents a unique opportunity to study the impact of gene duplication on NC evolution in multiple evolutionary contexts, including comparisons between non-teleost fishes (i.e. pre-WGD), multiple teleost fishes and other vertebrates.

For example, the differences in expression patterns between *sox10a* and *sox10b* in cichlids are more pronounced than those observed in medaka, particularly with respect to the apparent absence of *sox10a* expression in the extraembryonic tissues in the latter species (Nagao et al., 2018; Tsunogai et al., 2021). The acquisition of novel and divergent expression domains in cichlids suggests the presence of new regulatory elements, warranting future studies to investigate the expression and regulation of *sox10* paralogs in this clade and teleosts on a broader scale.

Moreover, we showed that *sox10a* knockout mutant is associated with aberrant craniofacial skeletal development phenotypes in *A. calliptera*, consistent with the emergence of functional divergence between *sox10a and sox10b* in cichlid fishes and other teleosts. The observed cartilage malformations in mosaic *sox10a*-CRISPR cichlids contrast with the effects of complete knockout mutations of both its orthologs *sox10a* and *sox10b* in medaka and *sox10b* in zebrafish (Kelsh and Eisen, 2000; Nagao et al., 2018). Functional analyses in medaka (Nagao et al., 2018) and zebrafish (Kelsh and Eisen, 2000) suggest that in these lineages *sox10* paralog(s) have retained ancestral functions in pigmentation, neuron and glial cell development, consistent with observations in other vertebrate lineages **(Fig. 7, Supplementary Table 5)**. Notably, medaka *sox10a* and *sox10b* have partially redundant functions in development of pigment cells , in agreement with the most common scenario of subfunctionalization of gene duplicates, also reported for *sox9* paralog in zebrafish (Nagao et al., 2018; Yan et al., 2005). In these teleosts, pigmentation was severely reduced, but cartilage development remained unaffected, similar to other vertebrates (Honoré et al., 2003; Kapur, 1999). The chondrogenesis aberration seen in *sox10a* mutants is more reminiscent of the effects of *sox9a* homozygous mutation in Nile tilapia (Li et al., 2023) and zebrafish (Yan et al., 2005), as well as *sox9* in mice (Wagner et al., 1994) **(Fig. 7, Supplementary Table 5)**.

The expression and functional data combined thus suggest a different partitioning of functions among soxE family genes (specifically *sox9* and *sox10*) in cichlids compared to other teleosts. We posit that *sox10a* acquired an essential role in chondrogenesis in the cichlid lineage, a function performed by *sox9* genes in other teleosts (Cresko et al., 2003; Nagao et al., 2018; Yan et al., 2005), possibly overlapping with the role of *sox9a* in cichlids (Li et al., 2023) **(Fig. 7)**. In contrast, cichlid *sox10b* appears to have retained its function in pigmentation development, akin to its orthologs in other teleost lineages. The subtle, yet significant, pigmentation phenotypes of *sox10b*-CRISPR fish could be due to the mosaic nature of the induced mutations. It remains to be determined whether potential differences in head pigmentation in severe *sox10a* mutants reflect some functional overlap between paralogs or indirect effects of the cranial cartilage malformation on pigment cell development. Similarly, the striking eye and brain size reduction in *sox10a* mutants

suggests that chondrogenesis of the craniofacial skeleton might also influence ocular and brain development. Together with observed expression differences of *prdm1a (sox10* regulator) at the anterior neural plate border, these findings may also implicate early variation in cranial placode (developmental precursors of sensory systems) development between species. More comprehensive analyses of NC development in *sox10* mutants, along with examination of non-NC-derived lineages expressing *sox10b,* will be required to better understand the functions of these genes during cichlid embryogenesis. Future studies in a wider range of species will be necessary to assess the extent of functional overlap among other *soxE* family members and to delineate the roles of *sox10a* and *sox9* paralogs in cichlid chondrogenesis, cranial placode and ocular development.

Conclusion

Functional divergence of duplicated genes is widely recognised for its role in the evolution of morphological diversity, including the expansion of the pigmentation pathway in teleosts (Braasch et al., 2009). In contrast, neofunctionalization is a rare occurrence in genetic evolution, with subfunctionalization (partitioning of ancestral functions between paralogs) being the most common scenario, allowing for developmental fine-tuning (Force et al., 1999). Our findings in cichlids, however, reveal that *sox10a* has acquired a novel function in chondrogenesis, which has not been previously reported in any other teleost clade. This shows that these paralogs have followed divergent evolutionary fates throughout teleost evolution.

We further hypothesise that, while cichlid *sox10b* retained its ancestral function, the neofunctionalization of *sox10a* may have contributed to the rewiring of the NC genetic and developmental programmes. These programmes show remarkable differences between cichlid species, thus providing an opportunity and genetic raw material for the remarkable diversification of the NC-derived phenotypes in cichlids. Altogether, *sox10* paralogs and their variable evolutionary fates are an ideal system for studying the evolution of NC gene regulatory networks across multiple evolutionary scales, within cichlid radiations, and among teleosts and vertebrate species.

Materials and Methods

Animal husbandry and embryo culture

Breeding stocks of *Astatotilapia calliptera* 'Mbaka' and *Rhamphochromis* sp. 'chilingali' were maintained under standardised conditions as previously described in (Marconi et al., 2023). Eggs used for RNA extractions and HCR *in situ* hybridisation experiments were collected from mouthbrooding females immediately after fertilisation and then reared individually in 1 mg/L of methylene blue (Sigma Aldrich) in water in 6-well plates (ThermoFisher Scientific) placed on an orbital shaker moving at slow speed at 27°C until needed. All experiments were conducted in compliance with the UK Home Office regulations.

Whole embryo bulk RNA sequencing

Sample acquisition

For each examined species, all samples were collected from the same egg clutch. Sampling covered the entire period of somitogenesis, which coincides with NC development, with samples taken at 3-hour intervals. At each time point, at least four embryos were dissected and placed individually into either 250 µl of pre-chilled Trizol (Ambion) and stored at -80°C until RNA extraction (at least overnight) or into 1 ml of 4% PFA in 1X PBS for overnight fixation at 4°C. Embryos preserved in 4% PFA were later rinsed twice in 1X PBS and stained with 10 nM DAPI in 70% glycerol in 1X PBS overnight at 4°C, protected from light. Following a wash in 1X PBS (10 min/wash, once), the embryos were mounted on microscopy slides (ThermoFisher) with Fluoromount G (Southern Biotech) and imaged with an Olympus FV3000 confocal microscope to confirm the developmental age (somite stage) of each sampled cohort.

RNA extraction

All procedures were conducted on ice, unless otherwise specified. Samples stored in Trizol were thawed from -80°C. For each sample, 100 mg of 0.1 mm zirconia/silica beads (Stratech) were added before homogenization using a TissueLyser II (Qiagen) for 120 seconds at 30 Hz. The

samples were then topped up to 1 ml with chilled Trizol and allowed to rest for 5 minutes. Next, 200 µL of chloroform (ThermoFisher Scientific) was added and the samples were vigorously shaken for 15 seconds, briefly vortexed and incubated at room temperature for 15 minutes. The samples were then centrifuged at 300 x*g* for 20 minutes at 4°C. The supernatant was carefully transferred to a fresh tube and further processed using the Direct-Zol RNA Microprep Kit (Zymo) according to the manufacturer's instructions. The quality and quantity of the extracted total RNA were assessed using Qubit (RNA HS assay, Agilent) and Tapestation (Agilent). Total RNA extracted from each embryo was submitted individually for sequencing, with quantities ranging from 135 ng to 1.3 µg per sample. All sequenced samples had eRIN values above 9.3.

NGS library preparation

All libraries were prepared, quality-controlled and sequenced by Novogene Corporation (China) using the Illumina NovaSeq 6000 platform to generate paired-end reads of 150 base pairs (bp). On average, 32.49 ± 2.5 Mio paired-end 150bp reads were generated per sample (**Supplementary Table 1**).

Adapter trimming and quality filtering

The adapter sequences in reads were removed, and low-quality sequences (Phred<20) were filtered out with TrimGalore (v0.6.6).

Mapping of RNAseq reads to reference genome

All RNAseq reads were mapped to the *Astatotilapia calliptera* genome assembly (fAstCal1.2 in Ensembl 105) using STAR v.2.7.1a (Dobin et al., 2013).

Gene expression quantification

The number of reads mapped to each gene in the reference genome was counted in STAR using the built-in HTSeq-count option (Anders et al., 2015). Gene counts were normalised using the median of ratios method in DESeq2 (Love et al., 2014) (v1.34.0). A Principal component analysis

(PCA) was applied to reduce the dimensionality of the dataset using the R command prcomp (R 4.2.0).

Differential gene expression (DE) analysis and gene annotation

DE analysis was performed on a gene count matrix using DESeq2 (Love et al., 2014) (v.1.34.0). Genes with mRNA counts < 10 per sample in each species were filtered out prior to analysis and technical replicates for each stage were collapsed. Heatmaps of scaled gene expression (*Z*-score per gene across all samples using mean DESeq2-normalised gene count per somite stage per species) were generated using pheatmap (v.1.0.12) and clusters of genes (unbiased complete linkage clustering) were identified and then plotted using ggplot2 (v.3.3.6).

Genes were annotated using the reference genome *A. calliptera* (fAst.cal 1.2) with biomaRt package (Durinck et al., 2009) (v.2.54.0). Significantly differentially expressed genes for each pairwise comparison were identified by filtering for log2 Fold Change above the absolute value of 0.585 (i.e. ≥ 1.5-fold difference in expression) and adjusted p-value < 0.05.

Gene ontology (GO) analysis

GO annotation and functional enrichment analysis were carried out using gProfiler2 (Kolberg et al., 2020). The full extent of the GO annotation, including the most specific GO terms for each gene or gene product, was used rather than the commonly used *GOslim* annotation, which comprises only a subset of the terms belonging to each parent domain, thus reflecting the broader biological categories. To focus the candidate search on genes involved in the NC development, a dataset comprising all DE genes with GO terms broadly associated with the developmental programme of the NC (including GO terms related to its development, migration and differentiation), development of pigmentation as well as craniofacial complex was compiled (**Supplementary Table 2**). These were then used to explore differences between species.

Transposable elements and repeat expression quantification

Transposable elements and repeats were predicted using RepeatModeler (v. 2.0.2 with LTRStruct parameter) and were then annotated using the RepeatModeler custom library in the *A. calliptera* genome using RepeatMasker (v.4.1.4; http://repeatmasker.org/). Only transposon elements (TE) were further analysed (simple repeats, tRNA, rRNA, scRNA and satellites were excluded from subsequent analyses). To quantify TE gene expression, TEcount (v.1.0.1) from the TEtranscript package(Jin et al., 2015) was used following STAR mapping with the following parameters --chimSegmentMin 10 --winAnchorMultimapNmax 200 --outFilterMultimapNmax 100. Finally, a DESeq2-normalised gene count matrix for 1609 different transcribed TEs was generated and PCA (centred and scaled) was produced with R (prcomp).

Population genomics

Detection of sites under positive selection

To identify recent signatures of positive selections, we conducted genome-wide scans to detect regions with unusually high local haplotype homozygosity and measured the extended haplotype homozygosity between the two populations (XP-EHH) using REHH (v.3.2.2)(Gautier and Vitalis, 2012). We utilised available VCF files containing biallelic SNPs from 43 samples of the Rhamphochromis genus (23.6±8.3x sequencing depth, mean±sd) and 45 randomly chosen *A. calliptera* 'Masoko Benthic' (16.02±1.63x) (**Supplementary Fig. 2a**; see Methods in https://github.com/tplinderoth/cichlids/tree/master/callset and (Munby et al., 2021)). Following the approach by (Ravinet et al., 2018), we calculated between-population extended haplotype homozygosity (xpEHH) using phased variants with a minor allele frequency (MAF) >0.05. In total, 551 significant individual sites (SNPs) were identified across all chromosomes (log[*P* value] ≥ 3), suggesting potential positive selection. Significant sites found within 50kbp (±50kbp) were grouped into 'islands' using bedTools (2.29.2), resulting in 154 regions, which comprised between 1 and 39 significant SNPs (3.6 SNPs on average) and varied in size between 1bp and 69Mb (mean=4.2kb). Genes in the closest proximity to these islands were identified using closestBed (bedTools). In total, 184 genes were linked to putative islands of selection (≥1 gene per island), of which 73 were DEGs (**Supplementary Table 4**; 48, 5 and 20 islands located in gene bodies, promoter and intergenic

regions, respectively). To identify enrichment in sites under potential positive selection within gene expression clusters, permutation tests were performed between the observed distribution of xpEHH p-values across candidate genes (25kbp regions upstream of TSS) for each gene expression cluster and the expected distribution (over chance). Expected values were calculated by randomly selecting xpEHH values (median values) across size-matched genic windows (1000x iterations; **Supplementary Fig. 2d**).

Phylogenetic reconstructions

The coding sequences of *soxE* gene family members were retrieved from Ensembl (108) and directly from genome assemblies from Parey at al. (2023) for inclusion in the phylogenetic analyses (**Supplementary Table 7**). Multiple sequence alignments of *soxE* gene family members were constructed using Clustal Omega (Sievers and Higgins, 2018). Next, the alignments were trimmed with TrimAl (Capella-Gutiérrez et al., 2009) and used to infer evolutionary relationships with the maximum likelihood method in IQ-TREE v.1.6.12 (Nguyen et al., 2015). In-built ModelFinder (Kalyaanamoorthy et al., 2017) was used to infer the best-fit substitution model (TN+F+I+G4) based on the Bayesian information criterion (BIC). The branch supports for maximum likelihood analyses were obtained using the ultrafast boot-strap (UBS)(Kalyaanamoorthy et al., 2017) with 1000 replicates. Phylogenetic tree shown in **Fig. 4** was visualised using iTOL v.6 (Letunic and Bork, 2019) and only non-significant bootstrap values are shown (≤75%).

Genome editing

gRNA design and synthesis

Targets for CRISPR/Cas9 editing were selected with the CHOPCHOP software online (http://chopchop.cbu.uib.no/) (Labun et al., 2019) using the *Astatotilapia burtoni* genome (AstBur1.0) as a reference. Sequence similarity searches with BLAST against the *A. calliptera* genome (AstCal1.2, Ensembl 108) were performed to confirm homology and test for off-target effects. Two sgRNAs targeting exon 1 were designed for *sox10b* (5'-CTCGTCGTCGGATTTGACGG-3' and 5'-

CGCGGATTCCCGCGGGGAA-3') and *sox10a* (5'-CGGTCAGTCAGGTGCTGGACGGG-3' and 5'-TCGTTTCCCGATCGGCATAA-3'), respectively, and purchased from Integrated DNA Technologies (ITD) as Alt-R CRISPR-Cas9 sgRNA (2 nmol).

Microinjection

Single-cell embryos of *Astatotilapia calliptera* 'Mbaka' were injected and maintained following the protocol described in (Clark et al., 2022). We were unable to perform microinjections in a similar m in *Rhamphochromis* due to their prolonged and unpredictable breeding behaviour, rendering it technically unfeasible.

Embryo imaging

Injected and control embryos were imaged daily until 12 days post-fertilisation using a Leica M205 stereoscope with a DFC7000T camera under reflected light darkfield. All specimens were positioned in 1% low melting point agarose (Promega) and anaesthetised with 0.02% MS-222 (Sigma-Aldrich) if required to immobilise during imaging.

Genotyping

Tissue samples were taken from specimens sacrificed by overdose of 0.5% MS-222 (Sigma-Aldrich) to extract genomic DNA using PCRBIO Rapid Extract Lysis Kit (PCRBiosystems). Fragments of 190-420bp surrounding predicted deletion sites were amplified using PCRBIO HS Taq Mix Red (PCRBiosystems) with an annealing temperature of 56°C using following primer pairs: *sox10b* 5'-CTGTCACCGGGTCATTCCTC-3' and 5'-GCGTTCATTGGCCTCTTCAC-3'; *sox10a* 5'-ATGGTCACTCACTGTCACCG-3' and 5'-CCTCCTCGATGAATGGCCTC-3'. Amplicons were purified with QIAquick PCR Purification Kit (Qiagen) before Sanger sequencing. All protocols were conducted following manufacturer's instructions. Sequence analysis to infer CRISPR edit sites was performed using the Synthego ICE CRISPR analysis tool (https://ice.synthego.com/).

Cartilage preparations

Embryos were stained for cartilage following the protocol of (Marconi et al., 2023) with following modifications: (1) all specimens were bleached to remove melanophore pigmentation using a solution of 0.05% hydrogen peroxide (Sigma) and 0.05% formamide (ThermoFisher) for 30-45 mins under light and (2) samples were cleared using first 50% then 70% glycerol:water solutions until complete sinking. Specimens were stored in 70% glycerol until imaging in 80% glycerol using a Leica M205 stereoscope with a DFC7000T camera under reflected light.

Whole mount *in situ* hybridisation by Hybridisation Chain Reaction (HCR)

Reagents

The HCR probes and hairpin sets were ordered from Molecular Instruments, whereas all required buffers were made following the instructions provided by the manufacturer. The HCR probe sets (14-20 pairs per gene) were designed using target gene template sequences retrieved from *A. calliptera* genome assembly (fAstCal1.2, Ensembl 108) (**Supplementary Table 8**). Each probe set was designed by the manufacturer to target transcript regions common to all splicing isoforms while minimising off-target effects.

Due to a very low genetic variation in the coding sequences between study species, we used the same probe sets per each target gene for both cichlid taxa examined. Probe specificity was verified by BLAST searches against the *A. calliptera* genome available on Ensembl (AstCal 1.2) as well as against unpublished *Rhamphochromis* sp. ‘chilingali’ assembly.

Protocol overview

Dissected embryos were fixed overnight at 4°C in 4% PFA in 1X PBS. The following day, they were rinsed twice in 1X PBST (1X PBS+0.01% Tween-20) without incubation and washed in 1X PBST (10 min/wash, twice) before a stepwise dehydration to 100% MetOH in pre-chilled solutions of 25%, 50% and 75% MetOH:PBST (10 mins/wash, at 4°C) and stored at -20°C until further analyses.

The embryos were then rehydrated from 100% MetOH to 1X PBST in reciprocal series (5 mins/wash, at 4°C), followed by washes in 1X PBST at room temperature (5 min/wash, twice).

mRNA *in situ* hybridization by chain reaction (HCR) was carried out according to the protocol of Andrews *et al.* (2020)(Andrews et al., 2021) for whole mount amphioxus embryos with the following modifications. (1) 2 pmol of each probe mixture (1 µL of 2 µM stock) per 100 µL of probe hybridization buffer were used and (2) 60 pmol of each fluorescently labelled hairpin (i.e. 2 µL of 3 µM stock) were applied per 100 µL of amplification buffer. Finally, the embryos were stained with 10nM DAPI in 70% glycerol in 1X PBS overnight at 4°C protected from light, washed in 1X SSCT (10 min/wash, twice) before mounting with Fluoromount G (Southern Biotech) on glass bottom dishes (Cellvis) without coverslip or microscopy slides (ThermoFisher Scientific) with #1.5 coverslips (Corning), depending on the size of the specimen. To prevent embryos from getting squashed when mounting on slides, thin strips of electrical tape were used as bridges to create space between the slide and a coverslip. Clear nail varnish was used to seal the edges of the slide and all samples were cured overnight at room temperature protected from light before imaging.

All *in situ* hybridization experiments were performed with multiple specimens from different clutches (at least 3 individuals per clutch, repeated at least once with specimens from alternative clutches) to fully characterise the expression patterns.

Confocal microscopy

Imaging of dissected and stained embryos was carried out with an inverted confocal microscope Olympus FV3000 at the Imaging facility of the Department of Zoology, University of Cambridge. Since the fluorescence intensity levels were only compared as relative signals within each sample (i.e., embryo), imaging was performed using optimal laser power and emission wavelength for each sample. Sequential acquisition mode was used to minimise the signal crosstalk across channels and all images were acquired at 1024x768 resolution and 12-bit depth.

Image processing for figure presentation

Confocal micrographs were stitched using the Olympus FV3000 software and processed with Fiji(Schindelin et al., 2012) to produce optical sections, collapse z-stacks and adjust image brightness and contrast where necessary following guidelines by Schmied & Jambor (2021)(Schmied and Jambor, 2021). The presented transverse optical sections are maximum projection across 5 adjacent slices. Auto-fluorescence and background noise in each image was removed by subtracting the average pixel intensity measured for each channel in regions of the embryo where no fluorescent signal was observed. Any remaining overexposed pixels were removed using the “Remove outliers” (radius = 0.5 pixel, threshold = 50) function in Fiji. Images presented as figures were smoothened by applying a Gaussian filter with σ = 0.5. Image look-up tables (LUTs) were taken from the BIOP plugin (https://imagej.net/plugins/biop-lookup-tables) and “ChrisLUTs” (Christophe Leterrier and Scott Harden; github.com/cleterrier/ChrisLUTs) package for Fiji/ImageJ. The images were processed for any background imperfections and assembled into figures in Adobe Photoshop 2023.

All plots and statistical analyses were made using R version 4.1.1. The MetBrewer package (https://github.com/BlakeRMills/MetBrewer) was used for colour palettes throughout this work.

REFERENCES

Abzhanov, A., Kuo, W.P., Hartmann, C., Grant, B.R., Grant, P.R., Tabin, C.J., 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. Nature 442, 563–567. https://doi.org/10.1038/nature04843

Abzhanov, A., Protas, M., Grant, B.R., Grant, P.R., Tabin, C.J., 2004. Bmp4 and Morphological Variation of Beaks in Darwin's Finches. Science 305, 1462–1465. https://doi.org/10.1126/science.1098095

Albertson, R.C., Kocher, T.D., 2006. Genetic and developmental basis of cichlid trophic diversity. Heredity 97, 211–221. https://doi.org/10.1038/sj.hdy.6800864

Albertson, R.C., Powder, K.E., Hu, Y., Coyle, K.P., Roberts, R.B., Parsons, K.J., 2014. Genetic basis of continuous variation in the levels and modular inheritance of pigmentation in cichlid fishes. Mol. Ecol. 23, 5135–5150. https://doi.org/10.1111/mec.12900

Almeida, M.V., Vernaz, G., Putman, A.L.K., Miska, E.A., 2022. Taming transposable elements in vertebrates: from epigenetic silencing to domestication. Trends Genet. 38, 529–553. https://doi.org/10.1016/j.tig.2022.02.009

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, T.G.R., Pönisch, W., Paluch, E.K., Steventon, B.J., Benito-Gutierrez, E., 2021. Single-cell morphometrics reveals ancestral principles of notochord development. Development 148, dev199430. https://doi.org/10.1242/dev.199430

Aoki, Y., Saint-Germain, N., Gyda, M., Magner-Fink, E., Lee, Y.-H., Credidio, C., Saint-Jeannet, J.-P., 2003. Sox10 regulates the development of neural crest-derived melanocytes in *Xenopus*. Dev. Biol. 259, 19–33. https://doi.org/10.1016/S0012-1606(03)00161-1

Betancur, P., Bronner-Fraser, M., Sauka-Spengler, T., 2010. Assembling Neural Crest Regulatory Circuits into a Gene Regulatory Network. Annu. Rev. Cell Dev. Biol. 26, 581–603. https://doi.org/10.1146/annurev.cellbio.042308.113245

Bolande, R.P., 1997. Neurocristopathy: its growth and development in 20 years. Pediatr. Pathol. Lab. Med. J. Soc. Pediatr. Pathol. Affil. Int. Paediatr. Pathol. Assoc. 17, 1–25.

Braasch, I., Brunet, F., Volff, J.-N., Schartl, M., 2009. Pigmentation Pathway Evolution after Whole-Genome Duplication in Fish. Genome Biol. Evol. 1, 479–493. https://doi.org/10.1093/gbe/evp050

Braasch, I., Schartl, M., Volff, J.-N., 2007. Evolution of pigment synthesis pathways by gene and genome duplication in fish. BMC Evol. Biol. 7, 74. https://doi.org/10.1186/1471-2148-7-74

Brandon, A.A., Almeida, D., Powder, K.E., 2023. Neural crest cells as a source of microevolutionary variation. Semin. Cell Dev. Biol., Special Issue: The genetic and developmental basis of phenotypic diversification 145, 42–51. https://doi.org/10.1016/j.semcdb.2022.06.001

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H.J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., Baroiller, J.-F., Barloy-Hubler, F., Berlin, A., Bloomquist, R., Carleton, K.L., Conte, M.A., D'Cotta, H., Eshel, O., Gaffney, L., Galibert, F., Gante, H.F., Gnerre, S., Greuter, L., Guyon, R., Haddad, N.S., Haerty, W., Harris, R.M., Hofmann, H.A., Hourlier, T., Hulata, G., Jaffe, D.B., Lara, M., Lee, A.P., MacCallum, I., Mwaiko, S., Nikaido, M., Nishihara, H., Ozouf-Costaz, C., Penman, D.J., Przybylski, D., Rakotomanga, M., Renn, S.C.P., Ribeiro, F.J., Ron, M., Salzburger, W., Sanchez-Pulido, L., Santos, M.E., Searle, S., Sharpe, T., Swofford, R., Tan, F.J., Williams, L., Young, S., Yin, S., Okada, N., Kocher, T.D., Miska, E.A., Lander, E.S., Venkatesh, B., Fernald, R.D., Meyer, A., Ponting, C.P., Streelman, J.T., Lindblad-Toh, K., Seehausen, O., Di Palma, F., 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513, 375–381. https://doi.org/10.1038/nature13726

Bronner, M.E., Simões-Costa, M., 2016. Chapter Seven - The Neural Crest Migrating into the Twenty-First Century, in: Wassarman, P.M. (Ed.), Current Topics in Developmental Biology, Essays on Developmental Biology, Part A. Academic Press, pp. 115–134. https://doi.org/10.1016/bs.ctdb.2015.12.003

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated

alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascenção, K., Rummel, C., Ovchinnikova, S., Mazin, P.V., Xenarios, I., Harshman, K., Mort, M., Cooper, D.N., Sandi, C., Soares, M.J., Ferreira, P.G., Afonso, S., Carneiro, M., Turner, J.M.A., VandeBerg, J.L., Fallahshahroudi, A., Jensen, P., Behr, R., Lisgo, S., Lindsay, S., Khaitovich, P., Huber, W., Baker, J., Anders, S., Zhang, Y.E., Kaessmann, H., 2019. Gene expression across mammalian organ development. Nature 571, 505–509. https://doi.org/10.1038/s41586-019-1338-5

Carney, T.J., Dutton, K.A., Greenhill, E., Delfino-Machín, M., Dufourcq, P., Blader, P., Kelsh, R.N., 2006. A direct role for Sox10 in specification of neural crest-derived sensory neurons. Development 133, 4619–4630. https://doi.org/10.1242/dev.02668

Cheng, Y.-C., Cheung, M., Abu-Elmagd, M.M., Orme, A., Scotting, P.J., 2000. Chick *Sox10*, a transcription factor expressed in both early neural crest cells and central nervous system. Dev. Brain Res. 121, 233–241. https://doi.org/10.1016/S0165-3806(00)00049-3

Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., Pierce, N.A., 2018. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Dev. Camb. Engl. 145, dev165753. https://doi.org/10.1242/dev.165753

Clark, B., Elkin, J., Marconi, A., Turner, G.F., Smith, A.M., Joyce, D., Miska, E.A., Juntti, S.A., Santos, M.E., 2022. Oca2 targeting using CRISPR/Cas9 in the Malawi cichlid Astatotilapia calliptera. R. Soc. Open Sci. 9, 220077. https://doi.org/10.1098/rsos.220077

Cresko, W.A., Yan, Y.-L., Baltrus, D.A., Amores, A., Singer, A., Rodríguez-Marí, A., Postlethwait, J.H., 2003. Genome duplication, subfunction partitioning, and lineage divergence: Sox9 in stickleback and zebrafish. Dev. Dyn. 228, 480–489. https://doi.org/10.1002/dvdy.10424

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Donoghue, P.C.J., Graham, A., Kelsh, R.N., 2008. The origin and evolution of the neural crest. BioEssays 30, 530–541. https://doi.org/10.1002/bies.20767

Drerup, C.M., Wiora, H.M., Topczewski, J., Morris, J.A., 2009. Disc1 regulates foxd3 and sox10 expression, affecting neural crest migration and differentiation. Development 136, 2623–2632. https://doi.org/10.1242/dev.030577

Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191. https://doi.org/10.1038/nprot.2009.97

Dutton, K.A., Pauliny, A., Lopes, S.S., Elworthy, S., Carney, T.J., Rauch, J., Geisler, R., Haffter, P., Kelsh, R.N., 2001. Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. Development 128, 4113–4125. https://doi.org/10.1242/dev.128.21.4113

Eames, B.F., Schneider, R.A., 2005. Quail-duck chimeras reveal spatiotemporal plasticity in molecular and histogenic programs of cranial feather development. Development 132, 1499–1509. https://doi.org/10.1242/dev.01719

Edgley, D.E., Genner, M.J., 2019. Adaptive Diversification of the Lateral Line System during Cichlid Fish Radiation. iScience 16, 1–11. https://doi.org/10.1016/j.isci.2019.05.016

El Taher, A., Böhne, A., Boileau, N., Ronco, F., Indermaur, A., Widmer, L., Salzburger, W., 2021. Gene expression dynamics during rapid organismal diversification in African cichlid fishes. Nat. Ecol. Evol. 5, 243–250. https://doi.org/10.1038/s41559-020-01354-3

Elkin, J., Martin, A., Courtier-Orgogozo, V., Santos, M.E., 2023. Analysis of the genetic loci of pigment pattern evolution in vertebrates. Biol. Rev. https://doi.org/10.1111/brv.12952

Fish, J.L., Sklar, R.S., Woronowicz, K.C., Schneider, R.A., 2014. Multiple developmental mechanisms regulate species-specific jaw size. Development 141, 674–684. https://doi.org/10.1242/dev.100107

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y., Postlethwait, J., 1999. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. Genetics 151, 1531–1545. https://doi.org/10.1093/genetics/151.4.1531

Gans, C., Northcutt, R.G., 1983. Neural Crest and the Origin of Vertebrates: A New Head. Science 220, 268–273. https://doi.org/10.1126/science.220.4594.268

Gautier, M., Vitalis, R., 2012. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28, 1176–1177. https://doi.org/10.1093/bioinformatics/bts115

Gebuijs, I.G.E., Raterman, S.T., Metz, J.R., Swanenberg, L., Zethof, J., Van den Bos, R., Carels, C.E.L., Wagener, F.A.D.T.G., Von den Hoff, J.W., 2019. Fgf8a mutation affects craniofacial development and skeletal gene expression in zebrafish larvae. Biol. Open 8, bio039834. https://doi.org/10.1242/bio.039834

Hernandez-Lagunas, L., Choi, I.F., Kaji, T., Simpson, P., Hershey, C., Zhou, Y., Zon, L., Mercola, M., Artinger, K.B., 2005. Zebrafish narrowminded disrupts the transcription factor prdm1 and is required for neural crest and sensory neuron specification. Dev. Biol. 278, 347–357. https://doi.org/10.1016/j.ydbio.2004.11.014

Honoré, S.M., Aybar, M.J., Mayor, R., 2003. Sox10 is required for the early development of the prospective neural crest in Xenopus embryos. Dev. Biol. 260, 79–96. https://doi.org/10.1016/S0012-1606(03)00247-1

Howard, A.G., IV, Baker, P.A., Ibarra-García-Padilla, R., Moore, J.A., Rivas, L.J., Tallman, J.J., Singleton, E.W., Westheimer, J.L., Corteguera, J.A., Uribe, R.A., 2021. An atlas of neural crest lineages along the posterior developing zebrafish at single-cell resolution. eLife 10, e60005. https://doi.org/10.7554/eLife.60005

Jheon, A.H., Schneider, R.A., 2009. The Cells that Fill the Bill: Neural Crest and the Evolution of Craniofacial Development. J. Dent. Res. 88, 12–21. https://doi.org/10.1177/0022034508327757

Jin, Y., Tam, O.H., Paniagua, E., Hammell, M., 2015. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics 31, 3593–3599. https://doi.org/10.1093/bioinformatics/btv422

Kague, E., Gallagher, M., Burke, S., Parsons, M., Franz-Odendaal, T., Fisher, S., 2012. Skeletogenic Fate of Zebrafish Cranial and Trunk Neural Crest. PLOS ONE 7, e47394. https://doi.org/10.1371/journal.pone.0047394

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285

Kapur, R.P., 1999. Early Death of Neural Crest Cells is Responsible for Total Enteric Aganglionosis in Sox10Dom/Sox10Dom Mouse Embryos. Pediatr. Dev. Pathol. 2, 559–569. https://doi.org/10.1007/s100249900162

Kelsh, R.N., 2006. Sorting out Sox10 functions in neural crest development. BioEssays 28, 788–798. https://doi.org/10.1002/bies.20445

Kelsh, R.N., Eisen, J.S., 2000. The zebrafish colourless gene regulates development of non-ectomesenchymal neural crest derivatives. Development 127, 515–525. https://doi.org/10.1242/dev.127.3.515

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., Peterson, H., 2020. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. https://doi.org/10.12688/f1000research.24956.2

Kratochwil, C.F., Geissler, L., Irisarri, I., Meyer, A., 2015. Molecular Evolution of the Neural Crest Regulatory Network in Ray-Finned Fish. Genome Biol. Evol. 7, 3033–3046. https://doi.org/10.1093/gbe/evv200

Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., Valen, E., 2019. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. 47, W171–W174. https://doi.org/10.1093/nar/gkz365

Lang, M., Miyake, T., Braasch, I., Tinnemore, D., Siegel, N., Salzburger, W., Amemiya, C.T., Meyer, A., 2006. A BAC library of the East African haplochromine cichlid fish Astatotilapia burtoni. J. Exp. Zoolog. B Mol. Dev. Evol. 306B, 35–44. https://doi.org/10.1002/jez.b.21068

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47, W256–W259. https://doi.org/10.1093/nar/gkz239

Li, X.-Y., Tang, Y.-H., Deng, W.-Y., Zheng, Y., Wang, L.-S., He, X., Xie, Q.-P., Li, Y.-Q., Deng, L.,

Wang, D.-S., Wei, L., 2023. Involvement of Sox9a in chondrogenesis and gonadal development in teleost Nile tilapia (Oreochromis niloticus). Zool. Res. 44, 729–731. https://doi.org/10.24272/j.issn.2095-8137.2023.024

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Marconi, A., Yang, C.Z., McKay, S., Santos, M.E., 2023. Morphological and temporal variation in early embryogenesis contributes to species divergence in Malawi cichlid fishes. Evol. Dev. 25, 170–193. https://doi.org/10.1111/ede.12429

Munby, H., Linderoth, T., Fischer, B., Du, M., Vernaz, G., Tyers, A.M., Ngatunga, B.P., Shechonge, A., Denise, H., McCarthy, S.A., Bista, I., Miska, E.A., Santos, M.E., Genner, M.J., Turner, G.F., Durbin, R., 2021. Differential use of multiple genetic sex determination systems in divergent ecomorphs of an African crater lake cichlid. https://doi.org/10.1101/2021.08.05.455235

Nagao, Y., Takada, H., Miyadai, M., Adachi, T., Seki, R., Kamei, Y., Hara, I., Taniguchi, Y., Naruse, K., Hibi, M., Kelsh, R.N., Hashimoto, H., 2018. Distinct interactions of Sox5 and Sox10 in fate specification of pigment cells in medaka and zebrafish. PLOS Genet. 14, e1007260. https://doi.org/10.1371/journal.pgen.1007260

Nasoori, A., 2020. Formation, structure, and function of extra-skeletal bones in mammals. Biol. Rev. 95, 986–1019. https://doi.org/10.1111/brv.12597

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300

Olesnicky, E., Hernandez-Lagunas, L., Artinger, K.B., 2010. prdm1a Regulates sox10 and islet1 in the development of neural crest and Rohon-Beard sensory neurons. genesis 48, 656–666. https://doi.org/10.1002/dvg.20673

Parey, E., Louis, A., Montfort, J., Bouchez, O., Roques, C., Iampietro, C., Lluch, J., Castinel, A., Donnadieu, C., Desvignes, T., Floi Bucao, C., Jouanno, E., Wen, M., Mejri, S., Dirks, R., Jansen, H., Henkel, C., Chen, W.-J., Zahm, M., Cabau, C., Klopp, C., Thompson, A.W., Robinson-Rechavi, M., Braasch, I., Lecointre, G., Bobe, J., Postlethwait, J.H., Berthelot, C., Roest Crollius, H., Guiguen, Y., 2023. Genome structures resolve the early diversification of teleost fishes. Science 379, 572–575. https://doi.org/10.1126/science.abq4257

Powder, K.E., Cousin, H., McLinden, G.P., Craig Albertson, R., 2014. A Nonsynonymous Mutation in the Transcriptional Regulator lbh Is Associated with Cichlid Craniofacial Adaptation and Neural Crest Cell Development. Mol. Biol. Evol. 31, 3113–3124. https://doi.org/10.1093/molbev/msu267

Powell, D.R., Hernandez-Lagunas, L., LaMonica, K., Artinger, K.B., 2013. Prdm1a directly activates foxd3 and tfap2a during zebrafish neural crest specification. Development 140, 3445–3455. https://doi.org/10.1242/dev.096164

Rahm, A.-K., Wiedmann, F., Gierten, J., Schmidt, C., Schweizer, P.A., Becker, R., Katus, H.A., Thomas, D., 2014. Functional characterization of zebrafish K2P18.1 (TRESK) two-pore-domain K+ channels. Naunyn. Schmiedebergs Arch. Pharmacol. 387, 291–300. https://doi.org/10.1007/s00210-013-0945-1

Ravinet, M., Elgvin, T.O., Trier, C., Aliabadian, M., Gavrilov, A., Sætre, G.-P., 2018. Signatures of human-commensalism in the house sparrow genome. Proc. R. Soc. B Biol. Sci. 285, 20181246. https://doi.org/10.1098/rspb.2018.1246

Rocha, M., Singh, N., Ahsan, K., Beiriger, A., Prince, V.E., 2020a. Neural crest development: insights from the zebrafish. Dev. Dyn. 249, 88–111. https://doi.org/10.1002/dvdy.122

Rocha, M., Singh, N., Ahsan, K., Beiriger, A., Prince, V.E., 2020b. Neural crest development: insights from the zebrafish. Dev. Dyn. 249, 88–111. https://doi.org/10.1002/dvdy.122

Salzburger, W., 2018. Understanding explosive diversification through cichlid fish genomics. Nat. Rev. Genet. 19, 705–717. https://doi.org/10.1038/s41576-018-0043-9

Sande-Melón, M., Marques, I.J., Galardi-Castilla, M., Langa, X., Pérez-López, M., Botos, M.-A., Sánchez-Iranzo, H., Guzmán-Martínez, G., Ferreira Francisco, D.M., Pavlinic, D., Benes, V., Bruggmann, R., Mercader, N., 2019. Adult sox10+ Cardiomyocytes Contribute to Myocardial

Regeneration in the Zebrafish. Cell Rep. 29, 1041-1054.e5. https://doi.org/10.1016/j.celrep.2019.09.041

Sanger, T.J., Mahler, D.L., Abzhanov, A., Losos, J.B., 2012. ROLES FOR MODULARITY AND CONSTRAINT IN THE EVOLUTION OF CRANIAL DIVERSITY AMONG ANOLIS LIZARDS. Evolution 66, 1525–1542. https://doi.org/10.1111/j.1558-5646.2011.01519.x

Santos, M.E., Lopes, J.F., Kratochwil, C.F., 2023. East African cichlid fishes. EvoDevo 14, 1. https://doi.org/10.1186/s13227-022-00205-5

Sauka-Spengler, T., Bronner-Fraser, M., 2008. A gene regulatory network orchestrates neural crest formation. Nat. Rev. Mol. Cell Biol. 9, 557–568. https://doi.org/10.1038/nrm2428

Schebesta, M., Serluca, F.C., 2009. olig1 expression identifies developing oligodendrocytes in zebrafish and requires hedgehog and notch signaling. Dev. Dyn. 238, 887–898. https://doi.org/10.1002/dvdy.21909

Schilling, T.F., Kimmel, C.B., 1994. Segment and cell type lineage restrictions during pharyngeal arch development in the zebrafish embryo. Development 120, 483–494. https://doi.org/10.1242/dev.120.3.483

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Schmied, C., Jambor, H.K., 2021. Effective image visualization for publications – a workflow using open access tools and concepts. https://doi.org/10.12688/f1000research.27140.2

Sievers, F., Higgins, D.G., 2018. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 27, 135–145. https://doi.org/10.1002/pro.3290

Simões-Costa, M., Bronner, M.E., 2015. Establishing neural crest identity: a gene regulatory recipe. Development 142, 242–257. https://doi.org/10.1242/dev.105445

Southard-Smith, E.M., Kos, L., Pavan, W.J., 1998. SOX10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. Nat. Genet. 18, 60–64. https://doi.org/10.1038/ng0198-60

Steventon, B., Mayor, R., Streit, A., 2014. Neural crest and placode interaction during the development of the cranial sensory system. Dev. Biol. 389, 28–38. https://doi.org/10.1016/j.ydbio.2014.01.021

Suzuki, N., Sekimoto, K., Hayashi, C., Mabuchi, Y., Nakamura, T., Akazawa, C., 2017. Differentiation of Oligodendrocyte Precursor Cells from Sox10-Venus Mice to Oligodendrocytes and Astrocytes. Sci. Rep. 7, 14133. https://doi.org/10.1038/s41598-017-14207-0

Takamiya, M., Stegmaier, J., Kobitski, A.Y., Schott, B., Weger, B.D., Margariti, D., Delgado, A.R.C., Gourain, V., Scherr, T., Yang, L., Sorge, S., Otte, J.C., Hartmann, V., Wezel, J. van, Stotzka, R., Reinhard, T., Schlunck, G., Dickmeis, T., Rastegar, S., Mikut, R., Nienhaus, G.U., Strähle, U., 2020. Pax6 organizes the anterior eye segment by guiding two distinct neural crest waves. PLOS Genet. 16, e1008774. https://doi.org/10.1371/journal.pgen.1008774

Tsunogai, Y., Miyadai, M., Nagao, Y., Sugiwaka, K., Kelsh, R.N., Hibi, M., Hashimoto, H., 2021. Contribution of sox9b to pigment cell formation in medaka fish. Dev. Growth Differ. 63, 516–522. https://doi.org/10.1111/dgd.12760

Tucker, A.S., Lumsden, A., 2004. Neural crest cells provide species-specific patterning information in the developing branchial skeleton. Evol. Dev. 6, 32–40. https://doi.org/10.1111/j.1525-142X.2004.04004.x

Turner, G.F., 2007. Adaptive radiation of cichlid fish. Curr. Biol. 17, R827–R831. https://doi.org/10.1016/j.cub.2007.07.026

Voldoire, E., Brunet, F., Naville, M., Volff, J.-N., Galiana, D., 2017. Expansion by whole genome duplication and evolution of the sox gene family in teleost fish. PLOS ONE 12, e0180936. https://doi.org/10.1371/journal.pone.0180936

Volff, J.-N., 2005. Genome evolution and biodiversity in teleost fish. Heredity 94, 280–294. https://doi.org/10.1038/sj.hdy.6800635

Wada, N., Javidan, Y., Nelson, S., Carney, T.J., Kelsh, R.N., Schilling, T.F., 2005. Hedgehog signaling is required for cranial neural crest morphogenesis and chondrogenesis at the

midline in the zebrafish skull. Development 132, 3977–3988. https://doi.org/10.1242/dev.01943

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Hustert, E., Wolf, U., Tommerup, N., Schempp, W., Scherer, G., 1994. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell 79, 1111–1120. https://doi.org/10.1016/0092-8674(94)90041-8

Weiner, A. m. j., Allende, M. l., Becker, T. s., Calcaterra, N.B., 2007. CNBP mediates neural crest cell expansion by controlling cell proliferation and cell survival during rostral head development. J. Cell. Biochem. 102, 1553–1570. https://doi.org/10.1002/jcb.21380

Weiner, A.M.J., Scampoli, N.L., Steeman, T.J., Dooley, C.M., Busch-Nentwich, E.M., Kelsh, R.N., Calcaterra, N.B., 2019. Dicer1 is required for pigment cell and craniofacial development in zebrafish. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 1862, 472–485. https://doi.org/10.1016/j.bbagrm.2019.02.005

Wells, J.N., Feschotte, C., 2020. A Field Guide to Eukaryotic Transposable Elements. Annu. Rev. Genet. 54, 539–561. https://doi.org/10.1146/annurev-genet-040620-022145

Woltering, J.M., Holzem, M., Schneider, R.F., Nanos, V., Meyer, A., 2018. The skeletal ontogeny of Astatotilapia burtoni – a direct-developing model system for the evolution and development of the teleost body plan. BMC Dev. Biol. 18, 8. https://doi.org/10.1186/s12861-018-0166-4

Yan, Y.-L., Willoughby, J., Liu, D., Crump, J.G., Wilson, C., Miller, C.T., Singer, A., Kimmel, C., Westerfield, M., Postlethwait, J.H., 2005. A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. Development 132, 1069–1083. https://doi.org/10.1242/dev.01674

ACKNOWLEDGMENTS

AM was supported by Wellcome Trust (WT) PhD Programme in Developmental Mechanisms ((215223/Z/19/Z), EMS is supported by a Natural Environment Research Council (NERC) Independent Research Fellowship (NE/R01504X/1) and GV is supported by an EMBO fellowship (ALTF 339-2022). The authors thank the Department of Zoology Imaging Facility for assistance with microscopy, and Dr. H. Svardal (University of Antwerp), Dr. G. Turner (Bangor University), Dr. D. Joyce and A. Smith (University of Hull) for the kind gifts of fish stocks and samples. We thank Dr. Ben Steventon and the Santos lab for critical comments on the manuscript. For open access, the authors have applied a CC BY public copyright licence to any author-accepted manuscript version arising from this submission.

AUTHORS' CONTRIBUTIONS

AM, GV and EMS devised the study and analyses. MJN collected wild specimens to establish laboratory stocks. RD contributed to whole-genome data collection. AM collected and performed RNA extractions on the samples. AM and GV performed data analyses on sequencing data. AM and AK collected embryonic material, performed HCR staining and imaging. AM performed CRISPR/Cas9 mutagenesis, embryo imaging, phenotyping, cartilage staining and DNA collection. AM, GV and EMS wrote and revised the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

INCLUSION & ETHICS

The Malawi cichlid samples (*Rhamphochromis* species) were collected ethically under prescribed permits, and the results and data are published under an Access and Benefit Sharing agreement with the Government of Malawi. We acknowledge the contributions of the Malawi Department of Fisheries and the Government of Malawi for their assistance in the collection of samples and the generation of data and results.

DATA AVAILABILITY

The raw RNA sequencing reads have been deposited in the Gene Expression Omnibus and will be made public upon publication. This data is made available on an open access basis for research use only. Any person who wishes to use this data for any form of commercial purpose must first enter into a commercial licensing and benefit sharing arrangement with the Government of Malawi.

CODE AVAILABILITY

All scripts used to analyse the data can be accessed at: https://github.com/Santos-cichlids/neural-crest-sox10-paralogs-cichlids